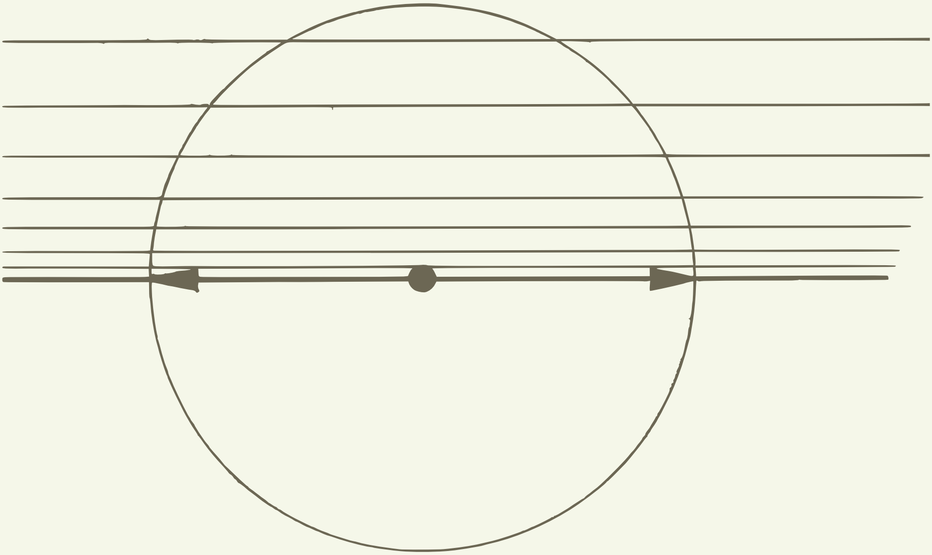
M. Postnikov

Lectures in Geometry

SEMESTER III

Smooth Manifolds



Mir
Publishers
Moscow

M. POSTNIKOV

LECTURES IN GEOMETRY

SEMESTER III

SMOOTH MANIFOLDS

Translated by Vladimir Shokurov

VOLUME 1

URSS
PUBLISHERS
MOSCOW

М. М. ПОСТНИКОВ
ГЛАДКИЕ МНОГООБРАЗИЯ
ТОМ 1
Москва, "Наука"

M. M. POSTNIKOV
SMOOTH MANIFOLDS
VOLUME 1
Moscow, "Nauka"

Translated by
Vladimir Shokurov

ISBN 5-88417-025-4
ISBN 5-88417-026-2 (vol. 1)

© М. М. Постников, 1987
© Translated by Vladimir Shokurov, 1989
© АО "УРСС", 1994

CONTENTS

VOLUME 1

Preface to the Russian Edition 10
Preface to the English Edition 16

Lecture 1 17
Simple lines in the plane. Giving lines by an equation. Whitney's theorem. Jordan curves. Smooth and regular curves. Nonparametrized curves. Natural parameter

Lecture 2. 36
Curves in the plane. Frenet formulas for a space curve. Projections of a curve onto the coordinate planes of the canonical frame. Frenet formulas for a curve in an *n*-dimensional space. The existence and uniqueness of a curve with given curvature

Lecture 3 51
Elementary surfaces and their parametrizations. Examples of surfaces. Tangent plane and tangent subspace. Smooth mappings of surfaces and their differentials. Diffeomorphisms of surfaces. The first quadratic form of a surface. Isometries. Beltrami's first differential parameter. Examples of computation of first quadratic forms. Developable surfaces

Lecture 4 80
Normal vector. Surface as the graph of a function. Normal sections. The second quadratic form of a surface. The Dupin indicatrix. Principal, total and mean curvatures. The second quadratic form of a graph. Ruled surfaces of zero curvature. Surfaces of revolution

Lecture 5 99
Weingarten formulas. Coefficients of connection. The Gauss theorem. Explicit formula for Gaussian curvature. The necessary and sufficient conditions of isometry. Surfaces of constant curvature

Lecture 6 110
Introductory remarks. Open subsets of the space R^n and their

diffeomorphisms. Charts and atlases. Maximal atlases. Smooth manifolds. Examples of smooth manifolds

Lecture 7 126
Topology of a smooth manifold. Open submanifolds. Neighbourhoods and interior points. Homeomorphisms. The first axiom of countability and the property of being locally flat. The second axiom of countability. Non-Hausdorff manifolds. Smoothnesses of a topological space. Topological manifolds. Zero-dimensional manifolds. The category TOP. The category DIFF. Pullback of smoothness

Lecture 8 142
Topological invariance of the dimension of a manifold. The dimension and coverings. Compact spaces. Lebesgue lemma. The upper estimate of the dimension of compact subsets of a space R^n. The monotonicity property of the dimension. Closed sets. The monotonicity of the dimension and closed sets. Direct product of topological spaces. The compactness of the direct product of compact spaces

Lecture 9 156
The drum theorem. The Brouwer fixed-point theorem. Cube separation theorem. Normal and completely normal spaces. Separation extension. The Lebesgue theorem on coverings of a cube. The lower estimate of the dimension of a cube

Lecture 10 170
Ordinals. Interval topology in sets of ordinals. Zero-dimensional spaces. Tychonoff's example. Tychonoff product of topological spaces. Filters. Centred sets of sets. Ultrafilters. Compactness criterions. The Tychonoff theorem

Lecture 11 186
Smoothness on an affine space. The manifold of matrices of a given rank. Stiefel manifolds. Matrix rows. The exponential of a matrix. The logarithm of a matrix. Orthogonal and J-orthogonal matrices. Matrix Lie groups. Groups of J-orthogonal matrices. Unitary and J-unitary matrices. Complex matrix Lie groups. Complex-analytic manifolds. Arcwise connected spaces. Connected spaces. The coincidence of connectedness

and arcwise connectedness for manifolds. Smooth and piecewise smooth paths. Connected manifolds failing to satisfy the second axiom of countability

Lecture 12 **210**
Vectors tangent to a smooth manifold. Derivatives of holomorphic functions. Tangent vectors to complex analytic manifolds. The differential of a smooth mapping. The chain rule. The gradient of a smooth function. The étale mapping theorem. The theorem on the change of local coordinates. Locally flat mappings

Lecture 13 **230**
Proof of the theorem on locally flat maps. Immersions and submersions. Submanifolds of a smooth manifold. A subspace tangent to a submanifold. Giving locally a submanifold. The uniqueness of the submanifold structure. The case of embedded submanifolds. The theorem on the inverse image of a regular value. Solutions of systems of equations. The group SL(n) as a submanifold

Lecture 14 **246**
The embedding theorem. Compact sets revisited. Urysohn functions. Proof of the embedding theorem. Manifolds satisfying the second axiom of countability. Scattered and meager sets. Null sets

Lecture 15 **260**
The Sard theorem. The analytic part of the proof of the Sard theorem. Direct product of manifolds. The manifold of tangent vectors. Proof of the Whitney theorem

VOLUME 2

Lecture 16 **274**
Tensors. Tensor fields. Vector fields and differentiations. Lie algebra of vector fields

Lecture 17 **291**
Integral curves of vector fields. Vector fields and flows. Transfer of vector fields using diffeomorphisms. The Lie

derivative of a tensor field

Lecture 18 304
Linear differential forms. Differential forms of arbitrary degree. Differential forms as functionals of vector fields. The inner product of a vector field and a differential form. Pullback of a differential form via a smooth mapping

Lecture 19 318
Exterior differential of a differential form. The Lie derivative of a differential form

Lecture 20 330
The de Rham complex and cohomology groups of a smooth manifold. The group H^0X. Poincaré lemma. The group H^1S^2. The group H^1S^1. Computing H^1S^1 using integrals. The group H^2S^2. Groups H^1S^n for $n \geq 2$. Groups $H^mS^n, m<n$.Groups H^nS^n

Lecture 21 358
Simplicial schemes and their geometric realizations. Cohomology groups of simplicial schemes. The double complex of a covering. Cohomology groups of a double complex. Augmented double complexes. Boundary homomorphisms. Acyclic complexes. Row acyclicity for $p=0$

Lecture 22 374
Row acyclicity of the double complex of a numerable covering. Row acyclicity of the double complex of a Leray covering. The de Rham-Leray theorem. Generalization. Groups E_2^{pq}. Groups F^{pq}. A group adjoined to a graded group with filtration

Lecture 23 389
Groups E_r^{pq}. Spectral sequences. Spectral sequence of a double complex. Spectral sequence of a covering

Lecture 24 405
Compactly exhaustible and paracompact topological spaces. Paracompact manifolds. Integrals in R^n. Cubable sets and

densities in arbitrary manifolds. Integration of densities

Lecture 25 420
Orientable manifolds. Integration of forms. Poincaré lemma for finite forms. The group $H^n_{fin}X$. An orientable manifold

Lecture 26 436
The degree of a smooth proper mapping. The algebraic number of inverse images of a regular value. Invariance of the degree under smooth homotopies. Proof of the drum theorem. Invariance of the degree under any homotopies

Lecture 27 448
Domains with regular boundaries. Stokes' theorem. Gauss-Ostrogradskii, Green's and Newton-Leibniz formulas. Manifolds with border sets. Interior and boundary points. Embedded ∂-submanifolds. Stokes' theorem for manifolds with border set and for ∂-submanifolds. Stokes' theorem for surface integrals. Stokes' theorem for singular submanifolds. Line integrals of the second kind

Lecture 28 468
Operators of vector analysis. Consequences of the identity $d \circ d = 0$. Consequences of differentiation formulas for products. The Laplacian and the Beltrami operator. The flow of a vector field. The Gauss-Ostrogradskii formulas for divergence and Green's formulas. Convergence as the density of sources. The Stokes formula for circulation. The Gauss-Ostrogradskii formula for rotation. The generalized Gauss-Ostrogradskii formula

Lecture 29 487
Periods of differential forms. Singular simplexes, chains, cycles, and boundaries. Stokes' theorem for chain integrals. Singular homology groups. The de Rham theorem. Cohomology groups of a chain complex. Singular cohomology groups

Subject index 508

PREFACE TO THE RUSSIAN EDITION

Geometry has been and remains the Cinderella of the curriculum at the Moscow University's Mathematics and Mechanics faculty. Never once during the last fifty years has the curriculum contained a course on the foundations of geometry or algebraic curves or transformation groups or even projective geometry (if we do not count the scraps in the first-semester courses in analytic geometry which may only be given under special circumstances, and nobody cares when the lecturer curtails them up or even drops them altogether). A student might well graduate from the faculty—with honours!—with no idea about Lobachevskian geometry, Caley-Klein ideas in the foundations of geometry, or the properties of algebraic curves or Lie groups.

Some twelve years ago the overflow into the calculus course of geometric material due to the ever increasing implementation of geometric methods led to the creation in the second year of a new course with the *ad hoc* name "Smooth manifolds and differential geometry". This course was delivered at a rate of a lecture a week and it was hoped that the course would free lecturers from presenting the extraneous geometric material. The course, however, was not well thought-out, and the parallel courses in calculus and differential equations were not coordinated with it. As a result the lecturers on calculus did not derive any advantage, and ridiculous as it may seem, integration of the differential forms on manifolds and Stokes' formula were discussed twice in as much detail but from slightly different points of view, in the two concurrent courses.

Delivering the geometry course in the third semester did not allow the generalizing and unifying role played by geometric concepts in modern mathematics to be brought out since to do so it is necessary for the main analytic courses to have been covered.

These and various more particular considerations led to the transfer of geometry to the third year (the fifth and sixth semesters). It has immediately become clear that this also had disadvantages.

A necessary constituent part of any course in geometry is the theory of curves and surfaces in three-dimensional Euclidean space, which is important both in its own right and as a source of examples and analogues for Riemannian geometry and the geometry of affine connections. By the third year, this material is too elementary (by this time students have already acquired a knack and a taste for more complicated constructions and concepts) and for it to play its propaedeutic role one cannot pass too quickly onto Riemannian geometry.

It is clear that this theory must be presented no later than the third semester (or perhaps earlier, as suggested by me in the first Russian edition of Semester II of these Lectures, even in the second semester). Moreover, a third-year geometry course does not help lecturers presenting second-year analysis (which I am sure will soon lead to the abolition of the course in geometry in the third semester and may be, alas, to its ousting from the schedule of the curriculum).

The radical solution is, of course, to overhaul the traditional system of mathematical courses. However, since there is an acute struggle between the departments for hours and courses such a review, which will have to be carried out sooner or later, is at present not possible, and a temporary solution would be the return of the geometry course in the third or fourth semester with the presentation of integration topics clearly distributed between the courses in calculus and geometry, each passing on the baton to the other as it were.

The following distribution of topics is suggested. After the integral of functions over domains in $\mathbb{R}^n$ has been discussed in calculus, the geometry lecturers cover the integration of densities and forms on manifolds. Simultaneously the calculus lecturer illustrates the general theory by particular cases of line and surface integrals of the first (density) and the second (form) kind. During this time, the generalized Stokes theorem is discussed which in the calculus course is immediately rendered concrete in the form of the Green, Gauss-Ostrogradskii and Stokes formulas. This duet, in which the general melody sometimes drifts, sometimes merges, ends in the apotheosis of vector analysis with elements of potential theory where the course in calculus changes freely into the theory of multidimensional improper integrals and the geometry course into cohomology theory. All this, of course, requires close coordination between the lecturers, which is not easy to achieve.

This book has, like its predecessors*, grown out of lectures given at the mathematics and mechanics faculty of Moscow University in different years. It is not, however, a recording of any particular course, but is instead a realization of the proposed geometry syllabus for the third semester. It can, however, certainly be used as the text for the fifth semester course.

The textbook is intended as a normal course presented at two lectures a week. The number of lectures (29) arises because although the winter semester formally contains 18 weeks, in practice it is impossible to deliver more than 11 to 15 weeks of lectures. The course can be used, however, even if the curriculum assigns only one or one and a half lectures a week (11 to 15 and respectively 16 to 22 lectures).

To be able to estimate the time required for a syllabus, I have tried to make each lecture in the book correspond to a two-hour lecture. Repeating material from other courses and considering examples, in written form, requires much more time. This accounts for difference in the volume of the lectures and the unexpectedly large size of Lectures 3, 11, and 20.

---

* M. Postnikov. *Lectures in Geometry*: *Semester* I. *Analytic Geometry*. (Mir Publishers, Moscow, 1981), *Semester* II. *Linear Algebra and Differential Geometry*. Mir Publishers, Moscow, 1982). (Referred to as I and II in what follows.)

The book focusses on smooth manifolds, and general topological facts and ideas are not separately presented being interspersed in the text.

In recent years a rather strange view of smooth manifolds has become widespread, a view surprisingly shared by some respected and competent mathematicians. Since a smooth manifold can be regarded as the result of the natural attempt to generalize axiomatically the simple idea of a manifold as a subset of a Euclidean space defined by a system of functionally independent equations, it is argued that the generalization does not actually lead to new objects because of the Whitney embedding theorem and so manifolds should be defined as subsets of that kind and that the general concept of a manifold is just an example of an axiomatic construction which inevitably arises in following a concept to its conclusion but one which it is then better to forget. I cannot share this opinion because in practice—for example, in mechanics—manifolds tend to appear in an abstract form, unembedded in a Euclidean space, and their forced embedding (with great arbitrariness!) introduces an additional structure that is sometimes useful but often having no relevance to the crux of the matter. The adherents of the former opinion appeal to Poincaré, who has shared it. In fact Poincaré clearly understood the necessity of having a general concept of manifold and dwelt on pasting together the charts of an atlas. Referring to the extremes of axiomatization is also wrong, since in reality manifolds were not introduced as the result of "natural attempt to generalize the simple notion of a manifold given by equations" but as an answer to the need to clearly explicate the notion arising in mathematical investigation. A consistent execution of the same principles would throw mathematics a hundred years back, since from this point of view, for example, all linear algebra in its present form has no right to exist, being based as it is, on the concept of a vector space which could be said to "have arisen as a result of a natural attempt to generalize the simple idea of the space $\mathbb{R}^n$", (which is as false as it is for manifolds), whereas the isomorphism theorem shows that "the generalization does not actually lead to new objects" (which, though true, does not deprive the concept of a vector space of its value). In this book, therefore, manifolds are defined in the usual way, on the basis of an atlas, while subsets of Euclidean spaces only appear as examples.

The problems in this book are mainly quite trivial and intended exclusively for a reader to test himself. Some more difficult problems are given in small print. Auxiliary material on algebra or calculus is also given in small print.

The first five lectures are only indirectly related to the theory of smooth manifolds, mainly being devoted to elementary differential geometry. The theory of curves (Frenet formulas) is followed by the first and second quadratic forms of a surface, the Weingarten formulas are derived, and Gauss's theorem on the invariance of total curvature is proved. Everything not directly related to the Gauss theorem has been omitted (the Meusnier theorem and the Euler theorem, geodesics, asymptotic curves, lines of curvature, and the like). When delivering lectures in the second year this material had sometimes to be postponed until the middle of the semester (so as to satisfy the needs of the course on differential equations, by introducing the general theory of smooth manifolds as soon as possible). Although this did remove some repetitions (for example, it was then unnecessary to define the differential of a smooth mapping twice, first for surfaces and then in the general case), it was barely justified methodologically (as it links elementary differential geometry, which is local in nature, to the theory of manifolds).

The theory of manifolds begins in Lecture 6. The first ten lectures (from the sixth to the fifteenth) are devoted to basic geometric notions and theorems of the theory of manifolds. In the shorter 11-lecture course one may omit seven of these lectures, reducing the first five lectures to four and sacrificing Lectures 8, 9, and 10 (which treat in the main the topological theory of dimension and the Tychonoff theorems), and Lecture 10 in a 16-lecture course. The remaining lectures in this group (particularly those on the Sard and the Whitney theorems) must, in my opinion, be kept in the course under all circumstances.

An 11-lecture course actually stops here. It turns out to be, however, possible to save, by slightly reducing and condensing the presentation, about an hour and a half of lecture time in presenting some of the material in Lectures 16 and 17. As to the theory of differential forms (Lectures 18, 19, and 20), it must be put off in the shortened course until the next semester (or left to the calculus lectures).

In a 16-lecture course it is possible to finish the course with Lecture 20, which demonstrates various ways of computing de Rham cohomology groups using the example of a sphere. This means that the "integration" Lectures 24 to 29 are excluded from the course and their material is thus left to the calculus course.

In Lectures 21 to 23 an attempt is made to expound the theory of homologies and cohomologies (up to spectral sequences!) in a form suitable for the compulsory course. This is made possible by changing the generally accepted point of view and giving up the treatment of simplicial homology theory, which is alleged to be geometrically obvious. I am pleased to note that a similar approach, at a more advanced level, is accepted in *Differential Forms in Algebraic Topology* by R. Bott and L. W. Tu, which must be read by anyone who wants to become acquainted with the basic ideas and constructions of the classical homology theory in a bright and up-to-date presentation. When time is lacking it is possible to omit the second half of Lecture 22 and all of Lecture 23.

Finally, the concluding Lectures 24 to 29, which, if desired, can be partially interchanged with Lectures 21 to 23, deal with integration. Here the presentation is deliberately incomplete (for example, nothing is said about additive functions of a set), since these lectures reflect only part of the general picture, and omit what relates to calculus. Lecture 28 can be left entirely to the calculus lecturer. It is also possible to confine oneself to just one lecture, Lecture 29, which is virtually independent of the previous four lectures.

PREFACE TO THE ENGLISH EDITION

This book is actually Semester 3 of my *Lectures in Geometry*. It starts a new subject, however, and is therefore independent of the previous two semesters.

The book has two major features that distinguish it from other textbooks on elementary smooth manifold theory. Firstly, a lot of space is allotted to topological dimensional theory, the most geometry-oriented branch of general topology, and an acquaintance with it will bring joy to lovers of elegant mathematical constructions which provide deep insights. Secondly, I've ventured, in this elementary course, to present the basic notions of the theory of spectral sequences, a tool whose power and significance is becoming increasingly clear in current studies. This is done without first expounding general (co)homology theory.

It is hoped that the two topics will appeal to the interested English reader.

The symbol □ signifies the end of a proof of a theorem.

December 31, 1988 *M. Postnikov*

Lecture I

Simple lines in the plane · Giving lines by an equation · Whitney's theorem · Jordan curves · Smooth and regular curves · Nonparametrized curves · Natural parameter

There are several different approaches to explicate the idea of a line, which yield different results. In simple cases, however, all the approaches give virtually the same result.

Let us first consider lines in the plane.

A set Γ in the plane is said to be a *graph* if there is a system of (Euclidean or affine) x-, y-coordinates and a differentiable (alternatively, continuous) function $f\colon I \to \mathbb{R}$ defined on the (closed, half-open or open) interval I of the real axis $\mathbb{R}$ such that a point with coordinates x and y is in Γ if and only if $x \in I$ and $y = f(x)$. Intuitively, all graphs, are, of course, lines.

A point p_0 of a set C in the plane is said to be *simple* if there is an open disk U with centre at p_0 such that the intersection $U \cap C$ is a graph.

The set C is said to be *connected* if it cannot be divided into two sets having the property that each limiting point of one set does not belong to the other. (This graphically means that the set consists of one piece.)

The set C in the plane is said to be a *simple line* if it is connected and consists of only simple points.

Problem 1. Prove that any graph is connected (and hence is a simple line).

The various ways of explicating the notion of a line differ mainly by nonsimple points that are allowed. We shall avoid discussing these questions once and for ever by agreeing to consider only simple lines.

A simple line may (or may not) have *end points*. There may be at most two of them. A simple line with two end points (having the shape of a bent closed interval of the number line) is called *closed*, and that with one end point (having the shape of a bent half-open interval of the number line) is called *half-open*. A simple line without end points may have the shape of a bent open interval of the number line or that of a bent circle. In the former case it is called *open*, and in the latter case *closed*. (Thus the term "closed" as applied to simple lines has two meanings! This historically established ambiguity must be constantly kept in mind.)

The usual way of giving lines in the plane is to define them by equations of the form

$$F(x, y) = 0, \tag{1}$$

where x, y are coordinates (affine or rectangular) in the plane, and F is a function of x, y. [The statement that a set $\mathscr{L}$ is given by equation (1) implies by definition that a point p of the plane is in $\mathscr{L}$ if and only if its coordinates x, y satisfy equation (1). When F is treated as a function in the plane this means that $p \in \mathscr{L}$ if and only if $F(p) = 0$.]

To obtain lines (in the sense of an explicit definition) it is certainly necessary to impose on F certain conditions. First of all it is natural to require that F be *continuous*. [If discontinuous functions are allowed, then equations of the form (1) may define an arbitrary set A of points of the plane; it suffices to take as F the function $1 - \chi$, where χ is the so-called *characteristic function* of A equal to unity for the points of A and to zero outside A.]

Recall from the course in calculus that a point p of the plane (or generally of an arbitrary metric—Euclidean in particular—space) is said to be an *interior point* of A if there is $\varepsilon > 0$ such that the ε-neighbourhood of that point (an open ball—a disk in the plane—of radius ε with centre at p) is entirely in A. A set consisting of only interior points is called *open*. A set C is said to be *closed* if its complement is open or, equivalently, if for any convergent sequence of points $p_n \in C$ its limit $\lim p_n$ is also in C.

A subset of an **affine** (real and finite-dimensional) space $\mathcal{A}$ is said to be *closed* (*open*) if it is closed (open) with respect to some Euclidean metric on $\mathcal{A}$.

Problem 2. Show that if a subset of $\mathcal{A}$ is closed (open) with respect to one Euclidean metric on $\mathcal{A}$, then it is closed (open) with respect to any other as well.

Remark 1. A closed simple line in the plane is a closed (and bounded) set (for each of the two meanings of the term "closed line"). On the contrary, none of the simple lines—the open one included—is an open set. Moreover, there are open lines (for example, the graph of a tangent) which are closed sets (necessarily unbounded).

Remark 2. It can be easily seen that *none of the simple lines has a single interior point.* Therefore Cantor suggested to consider arbitrary closed sets without interior points as lines in the plane. This definition offers some advantages but it is too general for most mathematical theories (while failing to cover, say, open simple lines).

It is obvious that for every continuous function F on a metric space $\mathcal{X}$ the set of all points in which F is zero is closed. This means that *an equation of the form* (1) *with a continuous function F can specify only closed sets of the plane.*

Problem 3. Show that, conversely, for any closed set C of $\mathcal{X}$ there is a continuous function F on $\mathcal{X}$ such that $p \in C$ if and only if $F(p) = 0$. [**Hint.** Consider on $\mathcal{X}$ the function

$$F(p) = \inf_{q \in C} \rho(p, q), \quad p \in \mathcal{X},$$

a distance from p to C; here ρ is a metric in $\mathcal{X}$.]

A point of a set (1) is said to be *nonsingular* if at this point both partial derivatives

$$\frac{\partial F}{\partial x} \text{ and } \frac{\partial F}{\partial y}$$

of F exist, are continuous, and at least one of them is nonzero. The other points of set (1) are called *singular points.*

The **implicit-function theorem** known from calculus states that in the neighbourhood of any nonsingular point each set (1) is a graph, i.e. *every nonsingular point is a simple point*. (The converse is false. For example, when $F(x, y) = x(x^2 + y^2)$ the set (1) consists of the points of the axis of ordinates $x = 0$ and hence all its points are simple. At the same time the point (0, 0) is its singular point.)

It follows that the *set of all nonsingular points of every set* (1) *is a union of simple lines* (joining, in general, at singular points). If therefore the number of singular points is not very large, for example, it is finite, then the set of form (1) corresponds quite well to an intuitive idea of lines. (And it is quite appropriate to call them so.) If, however, there are many singular points, then the situation is quite different. Namely, as was shown by the American mathematician Whitney, an equation of the form (1) with an infinitely differentiable function F can define any closed subset of the plane.

Whitney's theorem is related to an arbitrary finite-dimensional point affine space $\mathcal{A}$. Every function F on that space can be regarded, on choosing a reference point O, as a function on the associated vector space $\mathcal{V}$ and hence, after choosing in $\mathcal{V}$ a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, as a function on a Euclidean space $\mathbb{R}^n$. A function F is said to be *smooth function of class* C^∞ or C^∞-function if, as a function on $\mathbb{R}^n$, it has continuous partial derivatives of all orders. (It is clear that if this condition is satisfied for one choice of the frame $O\mathbf{e}_1, \ldots \mathbf{e}_n$, then it is so for any other choice of it.)

Theorem 1 (Whitney's theorem). *For any closed subset C of an affine space $\mathcal{A}$ there is a C^∞-function F on $\mathcal{A}$ such that $p \in C$ if and only if $F(p) = 0$.*

The proof of Theorem 1 rests on the following lemma, which is of interest:

Lemma 1. *There is a C^∞-monotonic function $\alpha: \mathbb{R} \to \mathbb{R}$ such that*:

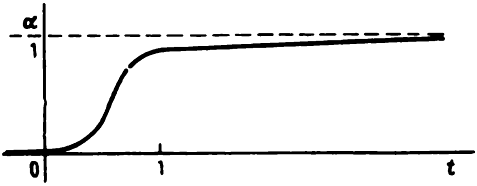
1° $0 \leqslant \alpha(t) < 1$ *for any* $t \in \mathbb{R}$,

2° $\alpha(t) = 0$ *if and only if* $t \leqslant 0$.

Proof. Put

$$\alpha(t) = \begin{cases} e^{-1/t} & \text{if } t > 0, \\ 0 & \text{if } t \leqslant 0. \end{cases} \tag{2}$$

Function (2) is clearly monotonic and has Properties 1° and 2°. Moreover, when $t \neq 0$ this function is obviously infinitely differentiable. We therefore must only prove that it is infinitely differentiable for $t = 0$ as well.



Graph of a function α

To this end recall that a function f defined in the neighbourhood of $t = 0$ is differentiable at that point if there are limits

$$\lim_{t \uparrow 0} \frac{f(t) - f(0)}{t}, \quad \lim_{t \downarrow 0} \frac{f(t) - f(0)}{t} \tag{3}$$

(the left and right derived numbers of f at $t = 0$) and if these limits are equal.

On the other hand, if f is differentiable in the neighbourhood of $t = 0$, except possibly for the point itself, and if there are limits

$$\lim_{t \uparrow 0} f'(t), \quad \lim_{t \downarrow 0} f'(t), \tag{4}$$

then, as it follows directly from the Lagrange theorem on finite increments, limits (3) exist and are equal to limits (4).

Since for all functions $f = \alpha^{(n)}$ the left-hand limits (4) obviously exist and are zero (for when $t < 0$ these functions are identically zero) it follows that to prove

Lemma 1 it suffices to establish that for any $n \geqslant 0$ the limit

$$\lim_{t \downarrow 0} \alpha^n(t) = \lim_{t \downarrow 0} (e^{-1/t})^{(n)}$$

exists and is equal to zero.

But it is easy to see that for any $n \geqslant 0$ there is a formula

$$(e^{-1/t})^{(n)} = e^{-1/t} p_n\left(\frac{1}{t}\right),$$

where $p_n = p_n(T)$ is a polynomial of degree $2n$. [This formula is true for $n = 0$; if it is true for some $n \geqslant 0$, then

$$\begin{aligned}(e^{-1/t})^{(n+1)} &= \left[e^{-1/t} p_n\left(\frac{1}{t}\right)\right]' \\ &= e^{-1/t}\left[\frac{1}{t^2} p_n\left(\frac{1}{t}\right) - \frac{1}{t^2} p_n'\left(\frac{1}{t}\right)\right] \\ &= e^{-1/t} p_{n+1}\left(\frac{1}{t}\right),\end{aligned}$$

where $p_{n+1}(T) = T^2 p_n(T) - T^2 p_n'(T)$ is a polynomial of degree $2n + 2$.]

Therefore

$$\lim_{t \downarrow 0} (e^{-1/t})^{(n)} = \lim_{t \to +\infty} \frac{p_n(t)}{e^t} = 0,$$

which was to be proved. ☐

Corollary 1. *For any closed interval* $[a, b]$ *of the axis* $\mathbb{R}$ *there is a* C^∞*-function* $\beta \colon \mathbb{R} \to \mathbb{R}$ *such that* $0 \leqslant \beta(t) \leqslant 1$ *for all* $t \in \mathbb{R}$ *and*

$$\beta(t) = \begin{cases} 1 & \text{if } t \leqslant a, \\ 0 & \text{if } t \geqslant b. \end{cases}$$

Proof. It suffices to put

$$\beta(t) = \frac{\alpha(b-t)}{\alpha(b-t) + \alpha(t-a)},$$

where α is the function in Lemma 1. ☐

Remark 3. We can construct C^∞-functions with a more complicated behaviour in a similar way. For example, for any numbers $a < c < d < b$ the formula

$$\lambda(t) = \frac{\alpha(B - |t - C|)}{\alpha(B - |t - C|) + \alpha(|t - D| - A)},$$

where

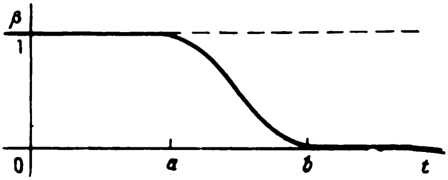
$$A = \frac{d-c}{2}, \quad B = \frac{b-a}{2}, \quad C = \frac{b+a}{2}, \quad D = \frac{d+c}{2},$$

defines a C^∞-function equal to zero outside $[a, b]$ and to unity on $[c, d]$. We shall need such a function in Lecture 15.

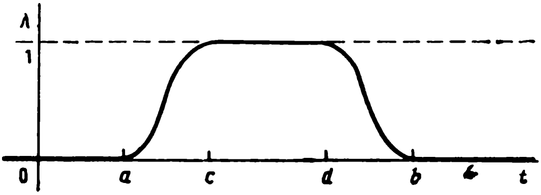
Let $\mathcal{V}$ be a Euclidean vector space.

Notation. For any $r > 0$ we shall denote by $\mathbf{B}_r^{\mathcal{V}}$ (or simply $\mathbf{B}_r$) a ball of radius r of $\mathcal{V}$ with centre at $\mathbf{0}$, i.e. the set of all vectors $\mathbf{x} \in \mathcal{V}$ for which $|\mathbf{x}| \leqslant r$. The corresponding *open ball* (a set of vectors $\mathbf{x} \in \mathcal{V}$ for $|\mathbf{x}| < r$) will be denoted by $\mathring{\mathbf{B}}_r^{\mathcal{V}}$.

For $\mathcal{V} = \mathbb{R}^n$ we shall write, as a rule, $\mathbf{B}_r^n$ rather than $\mathbf{B}_r^{\mathbb{R}^n}$.



Graph of a function β



Graph of a function λ

In a Euclidean point space $\mathcal{A}$ a ball of radius r with centre at p will be denoted by $\mathbf{B}_r^{\mathcal{A}}(p)$ (and open ball by $\mathring{\mathbf{B}}_r^{\mathcal{A}}(p)$).

These symbols will be constantly used throughout this course.

Corollary 2. *For any point p_0 of a Euclidean point space $\mathcal{A}$ and any $r > 0$ there is a function $f\colon \mathcal{A} \to \mathbb{R}$ such that $0 \leqslant f \leqslant 1$ on $\mathcal{A}$ and*

$$f(p) = \begin{cases} 1 \textit{ if and only if } p \in \mathbf{B}_r(p_0), \\ 0 \textit{ if and only if } p \notin \mathring{\mathbf{B}}_{2r}(p_0). \end{cases}$$

Proof. It suffices to put

$$f(p) = \beta(|\mathbf{x}|),$$

where $\mathbf{x} = \mathbf{p_0 p}$ is the radius vector of the point p counted off from p_0 and β is the function in Corollary 1, constructed for $[r, 2r]$. $\square$

On choosing in $\mathcal{A}$ a system of rectangular coordinates we call $p \in \mathcal{A}$ a *rational point* if all its coordinates are rational numbers. A ball $\mathring{\mathbf{B}}_r(p)$ will be called *rational* if its centre p and radius r are rational.

Lemma 2. *Every open set $U \subset \mathcal{A}$ is a union of a countable (or finite) number of rational balls, i.e. there are rational points $q_1, \ldots, q_m, \ldots$ and rational numbers $r_1, \ldots, r_m, \ldots$ such that*

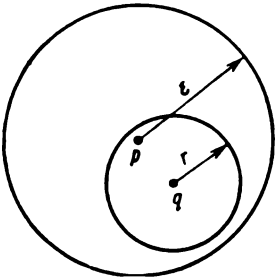
$$U = \bigcup_{m=1}^{\infty} \mathring{\mathbf{B}}_{r_m}(q_m). \tag{5}$$

Proof. Under the hypothesis, for any point $p \in U$ there is $\varepsilon > 0$ such that $\mathring{\mathbf{B}}_\varepsilon(p) \subset U$. Consider a rational ball $\mathring{\mathbf{B}}_r(q)$, where q is a rational point such that $\rho(p, q) < \varepsilon/2$ and r is a rational number such that $\rho(p, q) < r < \varepsilon/2$ (the existence of q and r is ensured by the fact that any real number may be approximated as much as desired by a rational one). Since $\rho(p, q) < r$, we have $p \in \mathring{\mathbf{B}}_r(q)$, and since

$$\rho(x, p) \leqslant \rho(x, q) + \rho(p, q) < 2r < \varepsilon$$

for any point $x \in \mathring{\mathbf{B}}_r(q)$, we have $\mathring{\mathbf{B}}_r(q) \subset \mathring{\mathbf{B}}_\varepsilon(p) \subset U$.

We thus see that any point $p \in U$ is contained in some rational ball $\mathring{\mathbf{B}}_r(q) \subset U$. This means that the set U is a union of rational balls of the form $\mathring{\mathbf{B}}_r(q)$ which are constructed for all possible points $p \in U$. But the set of all rational balls of a space $\mathcal{A}$ is clearly countable. The number of distinct balls of the form $\mathring{\mathbf{B}}_r(q)$ is therefore at most countable. Denoting them by $\mathring{\mathbf{B}}_{r_m}(q_m)$ we obtain expansion (5). ☐



We are now ready to prove Theorem 1.

Proof of Theorem 1. We may obviously assume without loss of generality that the affine space $\mathcal{A}$ under consideration is Euclidean and hence the complement $U = \mathcal{A} \setminus C$ of the set C allows a representation of the form (5). Let f_m be the function in Corollary 2 to Lemma 1 corresponding to q_m and $r_m/2$. This function (and hence each of its partial derivatives) is identically zero outside the compact (closed and bounded) set $\mathbf{B}_{r_m}(q_m)$. For any $k \geqslant 0$ therefore there is a number $c_m^k > 0$ such that the absolute value of each partial derivative of order k of f_m is at most c_m^k throughout $\mathcal{A}$. Let

$$c_m = \max(1, c_m^0, c_m^1, \ldots, c_m^m).$$

Consider a functional series

$$\sum_{m=1}^{\infty} \frac{f_m}{2^m c_m}. \tag{6}$$

Since by construction $c_m \geqslant 1$ and $f_m \leqslant 1$, this series is majorized by the numerical convergent series

$$\sum_{m=1}^{\infty} \frac{1}{2^m}. \tag{7}$$

Consequently, series (6) converges to some function $F: \mathcal{A} \to \mathbb{R}$. If $p \in C$, then $p \notin \mathbf{B}_{r_m}(q_m)$ for every $m \geqslant 1$ and hence $f_m(p) = 0$. But if $p \notin C$ (i.e. $p \in U$), then there is $m \geqslant 1$ such that $p \in \mathbf{B}_{r_m}(q_m)$. Then $f_m(p) \neq 0$ and hence $F(p) \neq 0$. Thus $F(p) = 0$ if and only if $p \in C$.

On the other hand, since for any $m \geqslant k$ each partial derivative of the m-th term of series (6) of order k obviously does not exceed the m-th term of series (7), on differentiating (6) k times we obtain a series all terms of which, except possibly for the first k terms, are also majorized by terms of series (7) and which is, consequently, uniformly convergent. According to the well-known series-differentiation theorem the sum of series (6) is infinitely differentiable (and each of its partial derivatives is the sum of a series made up of the corresponding partial derivatives of series (6)).

This completes the proof of Theorem 1. □

Another approach to the concept of a line, usually associated with the name of the French mathematician Jordan, is based on the concept of a line as a trajectory of a moving point. Lines in the sense of Jordan will be called curves.

According to Jordan, by a *curve* in an n-dimensional affine space $\mathcal{A}$ an arbitrary continuous mapping

$$\gamma: I \to \mathcal{A} \tag{8}$$

is meant, where I is some interval of the number line $\mathbb{R}$ (open, half-open or closed), i.e. after choosing the reference point in $\mathcal{A}$, a continuous vector function

$$\mathbf{r} = \mathbf{r}(t), \quad t \in I \tag{9}$$

which assumes values in the associated vector space $\mathcal{V}$.

In affine coordinates $x^1, \ldots x^n$ the Jordan curve (8) is given by continuous numerical functions

$$(10) \qquad x^1 = x^1(t), \ldots, x^n = x^n(t), \quad t \in I.$$

Equations (9) and (10) are called *parametric equations* of curve (8) (a *vector* and a *coordinate* one, respectively).

We stress that curves—in contrast to lines!—are not sets but mappings.

In practice, however, it is convenient to handle curves, at least from a terminological standpoint, as if they were sets. For example, for any $t_0 \in I$ a point $p_0 = \gamma(t_0)$ of $\mathcal{A}$ is called a *point of the curve* (8) *corresponding to the value of the parameter* t_0, and it is also said that *for* $t = t_0$ *curve* (8) *passes through* p_0. When the interval I is closed ($[a, b]$) points $\gamma(a)$ and $\gamma(b)$ are called the *end points* of the curve (8). Curve (8) is also said to join the point $\gamma(a)$ to the point $\gamma(b)$ and so on and so forth.

When $\gamma(a) = \gamma(b)$ curve (8) may be thought of as a continuous mapping of a circle. Such curves are called *closed.*

When it is required to stress the difference between a curve and the set of its points, the latter is called the *support* of the curve. Thus the support of curve (8) is nothing but the image $\gamma(I)$ of the interval I under mapping (8).

In general, the support of a curve may have the shape very far from the intuitive idea of a line. For example, it may have interior points or even fill a square, as shown by the example of a Peano curve.

Curve (8) is said to be *simple* if it is, first, an injective mapping $I \to \mathcal{A}$, i.e. $\gamma(t_1) = \gamma(t_2)$, $t_1, t_2 \in I$ if and only if $t_1 = t_2$, and, second, a reciprocal continuous mapping (also known as a *moneomorphic* mapping), i.e. such that if for a sequence $\{t_m\}$ of points of an interval I there is a point $\tau \in I$ such that $\lim \gamma(t_m) = \gamma(\tau)$, then $\{t_m\}$ converges (and $\lim t_m = \tau$). A closed curve

$$(11) \qquad \gamma\colon [a, b] \to \mathcal{A}, \quad \gamma(a) = \gamma(b)$$

is said to be *simple* if $\gamma(t_1) = \gamma(t_2)$ for $t_1 < t_2$ if and only if $t_1 = a$ and $t_2 = b$.

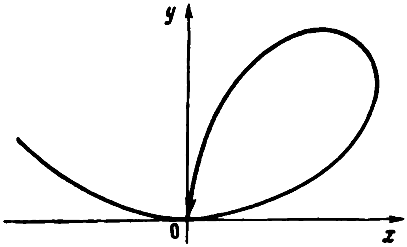
A typical example of an injective rather than moneomorphic mapping of an open interval into $\mathcal{A}$ is the curve

$$x=\frac{3t}{1+t^3},\quad y=\frac{3t^2}{1+t^3},\qquad -1<t<+\infty$$

(the "cut-off folium of Descartes").

Problem 4. Prove that for $I = [a, b]$ any injective mapping $I \to \mathcal{A}$ is moneomorphic.

The supports of simple curves are called *simple arcs.*



The cut-off folium of Descartes

In general, simple arcs already correspond to an intuitive idea of a line; in any event it follows from the theorem on the topological invariance of dimension (see Lecture 8 below) that they have no interior points (for $n > 1$). At the same time their structure is quite complicated.

Example 1. Let $x = x(t)$, $y = y(t)$, $0 \leqslant t \leqslant 1$ be parametric equations of the Peano curve in the plane. Then the equations

$$x = x(t),\quad y = y(t),\quad z = t,\quad 0 \leqslant t \leqslant 1$$

will give in space a simple arc whose projection onto the x, y plane is a square. Figuratively speaking, this means that a square area could be completely covered by a roof which is nevertheless a line rather than a surface!

Recall (from calculus) that a real function given on (a, b) is said to be a *smooth C^r-function*, where r is either a natural number or ∞, if it has continuous derivatives of all orders $\leqslant r$ (for $r = \infty$ this means by definition that there are continuous partial derivatives of all orders; see above).

In accordance with this we shall say that (8) given on $I = (a, b)$ is a *smooth C^r-curve* if all functions (10) are C^r-functions. Since the derivatives

$$x^i(t)' = \frac{dx^i(t)}{dt}, \quad i = 1, \ldots, n,$$

of functions (10) are the coordinates of a vector

$$\mathbf{r}'(t) = \lim_{h \to 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h}, \tag{12}$$

this condition means that continuous derivatives of all orders $\leqslant r$ for $r \neq \infty$ (continuous derivatives of all orders for $r = \infty$) exist for vector function (9).

In what follows we shall always consider the number r to be sufficiently large for all differentiations we need to have meaning, and no reference to the class C^r will be made as a rule.

When the interval I has end points (i.e. when either $I = [a, b]$, or $I = [a, b)$ or $I = (a, b]$, curve (8) is said to be smooth if it bounds some smooth curve (of the given class C^r) defined on some large interval $I' \supset I$.

Problem 5. Prove that this is equivalent to the fact that (for $r \neq \infty$) functions (10) have continuous derivatives of all orders $\leqslant r$ on (a, b) and the corresponding one-sided derivatives at points $t = a$ and/or $t = b$.

A closed curve (11) is said to be smooth if in addition the one-sided derivatives at $t = a$ and $t = b$ coincide.

Vector (12) is called a *tangent vector* to the smooth curve (11) at t. Somewhat loosely, it is also called a tangent vector at a *point* $\gamma(t)$. (For simple curves this terminology is quite valid, though.)

In Lecture 15 we prove Sard's theorem from which it follows in particular that the *support of a smooth curve has no interior points* (and is even the so-called set of measure zero). Since the projection of a smooth curve is obviously a smooth curve, it follows that the phenomenon described in Example 1 is impossible in the class of smooth curves.

It is interesting that a smooth curve may have breaks.

Example 2. A curve in the plane with equations

$$x = \alpha(t),\ y = \alpha(-t), \quad -\infty < t < +\infty, \tag{13}$$

where α is the function in Lemma 1, has a support which consists of two coordinate half-lines $x = 0$, $y \geqslant 0$ and $x \geqslant 0$, $y = 0$ meeting at right angles!

Curve (8) (or (1)) is called *regular at* t_0 if $\mathbf{r}'(t_0) \neq 0$. A curve regular at all points is called *regular*.

Notice that curve (13) is not regular at the break point $t = 0$. This is not accidental, since it is known from calculus that the *support of a simple curve* (8) *regular at* t_0 *has at* $\gamma(t_0)$ *a single tangent* ($\mathbf{r}'$, (t_0) serving as the direction vector).

The two curves

$$\gamma: I \to \mathcal{A}, \quad \gamma^*: I^* \to \mathcal{A}, \tag{14}$$

where I and I^* the intervals of the same type (both are closed, both are open or both are half-open), are called *equivalent* if there is a smooth (C^r-) function

$$\varphi : I^* \to I, \tag{15}$$

with a derivative nonzero everywhere, which maps I^* onto I and such that $\gamma^* = \gamma \circ \varphi$, i.e. such that

$$\gamma^*(t^*) = \gamma(\varphi(t^*)) \text{ for any } t^* \in I^*. \tag{16}$$

It is also said that the function φ brings about a *change of parameter* on the curve γ.

The equivalence classes of curves are called *nonparametrized curves*. To stress the difference between curves and nonparametrized curves the former are sometimes called *parametrized curves*.

A nonparametrized curve is said to be *smooth*, *simple* or *regular* if it is the equivalence class of a smooth, simple or regular parametrized curve. Since a curve equivalent to a smooth, simple or regular curve is obviously also smooth or respectively simple and regular, this definition is correct.

If curves (14) are connected by relation (16), where φ is in general an arbitrary function, then the supports of these curves coincide. *Equivalent curves* therefore *have the same support* (which is called the support of the corresponding nonparametrized curve), but the converse is in general false.

In the class of simple and regular curves, however, the situation is more satisfactory.

Proposition 1. *If both curves* (14) *are simple and regular, then they have the same support if and only if they are equivalent.*

Proof. If curves (14) are simple and have the same support, then a continuous (why?) mapping $\varphi = \gamma^{-1} \circ \gamma^*$ of I^* onto I is correctly defined. It is only necessary therefore to prove that the mapping φ is smooth and that its derivative is everywhere nonzero.

Let t_0^* be an arbitrary point on I^* and let $t_0 = \varphi(t_0^*)$. In that case, if $p_0 = \gamma^*(t_0^*)$, then $p_0 = \gamma(t_0)$ and the support of curves (14) has a single tangent at p_0. If $\mathbf{r} = \mathbf{r}(t)$ and $\mathbf{r} = \mathbf{r}^*(t^*)$ are vector parametric equations of curves (14), then $\mathbf{r}'(t_0)$ and $\mathbf{r}^{*\prime}(t_0^*)$ will be the direction vectors of the tangent. Therefore *these vectors are collinear.*

Since the curve γ is regular at t_0, we have $\mathbf{r}'(t_0) \neq 0$. Therefore if

$$x^1 = f^1(t), \ldots, x^n = f^n(t), \quad t \in I,$$

and

$$x^1 = g^1(t^*), \ldots, x^n = g^n(t^*), \quad t^* \in I^*,$$

are the coordinate parametric equations of curves (14), then it may be assumed without loss of generality that $\frac{df^1}{dt}(t_0) \neq 0$ and therefore, by virtue of $\mathbf{r}'(t_0)$ and $\mathbf{r}^{*\prime}(t_0^*)$ being collinear, that $\frac{dg^1}{dt}(t_0^*) \neq 0$.

But if $\frac{df^1}{dt}(t_0) \neq 0$, then by the **inverse function theorem** known from calculus f^1 is locally invertible, i.e. there exists an interval (a, b) on the x-axis which contains $x_0 = f^1(t_0)$, and a function $t = h(x)$ mapping that interval onto some other interval (α, β) of the axis t, containing t_0 (and contained in I), such that

$$h(f^1(t)) = t \text{ for any point } t \in (\alpha, \beta).$$

The smooth function h belongs to the same class C^r as the smooth function f^1 and its derivative is nonzero at x_0.

By construction $g^1(t_0^*) = f^1(t_0) = x_0 \in (a, b)$. Therefore, there is an interval (α^*, β^*) on t^*, containing t_0^* and contained in I^*, such that $g^1(t^*) \in (a, b)$ for any point $t^* \in (\alpha^*, \beta^*)$. Consequently, defined on (α^*, β^*) is a function

$$h \circ g^1 : t^* \mapsto h(g^1(t^*)) \tag{17}$$

which assumes values on (α, β). This function belongs to smooth functions of class C^r and has the property that its derivative is nonzero at t_0^*.

On the other hand, under the hypothesis

$$f^i(\varphi(t^*)) = g^i(t^*)$$

for any point $t^* \in I^*$ and any $i = 1, \ldots, n$; in particular, for $t^* \in (\alpha^*, \beta^*)$ and $i = 1$. Therefore $\varphi(t^*) = (h \circ g^1)(t^*)$ for $t^* \in (\alpha^*, \beta^*)$, i.e. function (17) is a restriction of the function φ to the interval (α^*, β^*). Hence φ belongs to the smoothness function of class C^r on (α^*, β^*) and its derivative is nonzero at t_0^*. Since t_0^* is an arbitrary point on I^*, and intervals of the form (α^*, β^*) cover the whole of this interval, this proves that φ is of class C^r on the entire interval I^* and that its derivative is nonzero everywhere on I^*.

This completes the proof of Proposition 1. □

Proposition 1 implies that simple regular curves are uniquely defined up to equivalence by their supports (and may therefore be identified with them). These supports are called *regular simple arcs*. A regular simple curve whose support is a regular simple arc $\mathcal{L}$ is called a *parametrization* of $\mathcal{L}$. As a rule we shall identify regular simple arcs with their parametrizations (considered up to equivalence).

Remark 4. Although generalizing Proposition 1 to arbitrary curves by more general changes of the parameter (for example, with vanishing derivatives) seems natural at first sight, alas it is one of the many far-fetched problems that are meaningless.

For $n = 2$ (in the plane) any graph is obviously a regular simple curve. Moreover, it can be shown (try to do it!) that *in the plane regular simple arcs are precisely simple lines*. Thus, in relation to simple lines all explicit definitions of the intuitive idea of a line lead to the same result.

Problem 6. Show that *in the plane every regular curve* (8) *is locally equivalent to a graph*, i. e. for any point $t_0 \in I$ there is its neighbourhood $(a, b) \subset I$ in $\mathbb{R}$ such that the curve $\gamma|_{(a,\ b)}$ is equivalent to a curve with equations of the form $x = t$, $y = f(t)$ (and is hence a regular simple curve).

Problem 7. Give an example which shows that a regular curve, which is an injective mapping, cannot nevertheless be a simple, curve (and may even have a whole interval of nonsimple points)e.

If $\mathcal{A}$ is a Euclidean space, then for any smooth curve (8) on I a function

$$t \mapsto |\mathbf{r}'(t)|, \quad t \in I,$$

is defined, which is the length of a tangent vector $\mathbf{r}'(t)$. This function is trivially continuous, and hence when $I = [a, b]$ there is an integral

$$s = \int_a^b |\mathbf{r}'(t)|\, dt$$

of this function over $[a, b]$. As shown in calculus, this integral is equal to the limit of the length of the broken lines which are a refinement of curve (8), i.e. to the *length of curve* (8).

Now let I be an arbitrary interval, and let $t_0 \in I$. Then the formula

$$s(t) = \int_{t_0}^{t} |\mathbf{r}'(t)|\, dt, \quad t \in I, \tag{18}$$

defines on I a smooth function mapping I onto some other interval J of the s-axis, which contains the point 0. This function is called the *arc length.* (Notice that it may assume negative values as well.)

If $s(t) = t - t_0$, then the parameter t is called natural. Allowing the inaccuracy generally accepted in calculus we may thus say that the *parameter t is natural if it is an arc length.*

The property of a parameter to be natural is equivalent to the identity $s'(t) = 1$. Since by definition $s'(t) =$

$| \mathbf{r}'(t) |$, we see therefore that the *parameter* t *on curve* (8) *is natural if and only if*

$$| \mathbf{r}'(t) | = 1 \textit{ for all } t \in I.$$

In particular, we see that a *curve referred to a natural parameter is trivially regular.*

Conversely, let curve (8) be regular. Then $| \mathbf{r}'(t) | > 0$ for all $t \in I$, and therefore function (18) is monotonic and an inverse function

$$\varphi: J \rightarrow I$$

is defined for it. The curve

$$\gamma_1 = \gamma \circ \varphi: J \rightarrow \mathcal{A}$$

is equivalent to the curve γ and we have for it

$$\mathbf{r}_1'(s) = \mathbf{r}'(t) \frac{d\varphi}{ds}(s) = \mathbf{r}'(t) \frac{1}{s'(t)},$$

where $t = \varphi(s)$. Since $s'(t) = | \mathbf{r}'(t) |$ it follows that

$$| \mathbf{r}_1'(s) | = 1 \text{ for all } s \in J,$$

i.e. that the parameter s on the curve γ_1 is natural.

We thus see that *every regular curve is equivalent to a curve referred to a natural parameter.*

Therefore, since we restrict ourselves to regular (and in addition, simple) curves, all curves under consideration may be, without loss of generality, referred to a natural parameter. It is important to bear in mind that *for a regular simple curve the natural parameter is defined up to a transformation of the form*

$$t \mapsto \pm t + t_0$$

(i.e. up to a starting point and the direction of the measurement chosen).

In what follows the natural parameter will, as a rule, be denoted by s.

Differentiation with respect to s will be denoted by a point:

$$\dot{\mathbf{r}}(s) = \frac{d\mathbf{r}(s)}{ds}, \ \ddot{\mathbf{r}}(s) = \frac{d^2\mathbf{r}(s)}{ds^2}, \ \ldots, \text{ etc.}$$

As we have already seen, the parameter s naturalness is equivalent to the identity

$$|\dot{\mathbf{r}}(s)| = 1 \qquad \text{for all } s.$$

In this connection it is good to have in mind the following lemma (in which s is, of course, not the natural parameter):

Lemma 3. *Let* $\mathbf{u} = \mathbf{u}(s)$ *be a vector-valued smooth function such that* $|\mathbf{u}(s)| = 1$ *for all* s. *Then*

$$\mathbf{u}(s)\dot{\mathbf{u}}(s) = 0 \qquad \textit{for every } s. \tag{19}$$

Proof. The equation $|\mathbf{u}(s)| = 1$ is equivalent to $\mathbf{u}(s)^2 = 1$. But it is easy to see that for a scalar (as well as for a vector) multiplication of vectors the usual product differentiation formula remains. In particular,

$$(\mathbf{u}^2)^{\cdot} = \dot{\mathbf{u}}\mathbf{u} + \mathbf{u}\dot{\mathbf{u}} = 2\mathbf{u}\dot{\mathbf{u}}.$$

Therefore, if $\mathbf{u}^2 = 1$, then $\mathbf{u}\dot{\mathbf{u}} = 0$. □

Corollary. *For any curve* $\mathbf{r} = \mathbf{r}(s)$ *referred to the natural parameter there is a formula*

$$\dot{\mathbf{r}}(s)\ddot{\mathbf{r}}(s) = 0 \qquad \textit{for every } s. \tag{20}$$

Lecture 2

Curves in the plane · Frenet formulas for a space curve · Projections of a curve onto the coordinate planes of the canonical frame · Frenet formulas for a curve in an n-dimensional space · The existence and uniqueness of a curve with given curvature

---

Let

$$\gamma\colon I \to \mathscr{A} \tag{1}$$

be an arbitrary regular and simple curve in an n-dimensional Euclidean space $\mathscr{A}$. As we know from the preceding lecture, curve (1) without loss of generality may be referred to the natural parameters s.

Let $\mathbf{r} = \mathbf{r}(s)$ be a vector parametric equation of curve (1) and let

$$\mathbf{t}(s) = \dot{\mathbf{r}}(s) \tag{2}$$

be its tangent vector. Since s is the natural parameter, vector (2) is a unit vector and the vector

$$\dot{\mathbf{t}}(s) = \ddot{\mathbf{r}}(s)$$

is orthogonal to it:

$$\mathbf{t}(s)\,\dot{\mathbf{t}}(s) = 0 \text{ for all } s.$$

Definition 1. The length $|\dot{\mathbf{t}}(s)|$ of a vector $\dot{\mathbf{t}}(s)$ is denoted by $k(s)$ (or simply k) and called the *curvature* of the curve (1) at s (or $\mathbf{r}(s)$).

For example, for a plane curve

$$k(s) = \sqrt{\ddot{x}^2(s) + \ddot{y}^2(s)},$$

where $x = x(s)$ and $y = y(s)$ are coordinate parametric equations of curve (1) in a Euclidean coordinate system x, y.

By the curvature of a curve referred to an arbtrary parameter is meant the curvature of an equivalent curve referred to the natural parameter. The formula for this curvature (which can be obtained by simple but rather cumbersome calculations using nothing but formulas for differentiation of functions) has, even for plane curves, quite a complicated form:

$$k = \left| \frac{x''y' - y''x'}{[(x')^2 + (y')^2]^{3/2}} \right|.$$

Obviously, the number k is the instantaneous velocity of rotation of a unit vector **t**. The greater the curvature of the curve, the greater the velocity. Hence the term "curvature".

In an oriented plane the so-called *relative curvature* k_{rel} equal to the curvature k can be considered when (for $k \neq 0$) vectors **t** and $\dot{\mathbf{t}}$ form a positively oriented basis for the plane, and to $-k$ otherwise. We shall need this curvature in Lecture 4.

Example 1. If

$$x = x_0 + sl, \quad y = y_0 + sm, \quad \text{where } l^2 + m^2 = 1,$$

i.e. if the curve in question is a straight line, then $\ddot{x} = 0$ and $\ddot{y} = 0$. Therefore $k = 0$ for all s, i.e. as was to be expected, the *curvature of a straight line is identically zero.*

Since linear functions are the only functions whose second derivative is identically zero, the converse is also true, i.e. a *curve whose curvature is identically zero is a straight line* (or its segment).

The point $\mathbf{r}_0 = \mathbf{r}(s_0)$ on a curve $\mathbf{r} = \mathbf{r}(s)$ is said to be the *point of rectification* if $k(s_0) = 0$.

Example 2. Parametric equations of a circle of radius R in the natural parameter s are, obviously, of the form

$$x = R\cos\frac{s}{R}, \quad y = R\sin\frac{s}{R}.$$

Since

$$\ddot{x} = -\frac{1}{R}\cos\frac{s}{R}, \quad \ddot{y} = -\frac{1}{R}\sin\frac{s}{R},$$

for a circle

$$k(s) = \frac{1}{R} \qquad \text{for all } s.$$

Thus the *curvature of a circle is constant and equal to the inverse of its radius.*

From the general Theorem 1 to be proved below the converse follows, namely a *curve with constant curvature is a circle* (or its arc).

If for some curve $k(s_0) \neq 0$, then a number $R(s_0) = \frac{1}{k(s_0)}$ is defined which is called the *radius of curvature* of the curve at the point in question.

A curve $\mathbf{r} = \mathbf{r}(s)$ is called a *generic curve* (*a curve of general type*) if there are no points of rectification on it, i.e. if $k(s) \neq 0$ for all s. At each point of such a curve the unit vector

$$\mathbf{n}(s) = \frac{\dot{\mathbf{t}}(s)}{k(s)}$$

is defined which is directed along the *normal* to the curve (i.e. along the straight line passing through the point of tangency at right angles to the tangent).

For any s the vectors $\mathbf{t}(s)$ and $\mathbf{n}(s)$ form an orthonormal basis which is called *Frenet's moving basis* for a given generic curve.

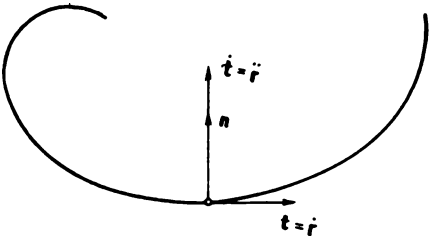
By definition

$$\dot{\mathbf{t}}(s) = k(s)\,\mathbf{n}(s).$$

Let us find a similar formula for $\dot{\mathbf{n}}(s)$. Let

$$\dot{\mathbf{n}}(s) = \alpha(s)\,\mathbf{t}(s) + \beta(s)\,\mathbf{n}(s)$$

be a decomposition of $\dot{\mathbf{n}}(s)$ in terms of the basis $\mathbf{t} = \mathbf{t}(s)$, $\mathbf{n} = \mathbf{n}(s)$. Since $\mathbf{tn} = 0$, we have $\dot{\mathbf{t}}\mathbf{n} + \mathbf{t}\dot{\mathbf{n}} = 0$ (we use again the fact that for a scalar product of vectors the formula for product differentiation is valid), and therefore $\alpha = \mathbf{t}\dot{\mathbf{n}} = -\dot{\mathbf{t}}\mathbf{n} = -k$. On the other hand, according to Lemma 2, Lecture 1, $\beta = \mathbf{n}\dot{\mathbf{n}} = 0$. This proves that for *any generic curve* formulas



The Frenet basis for a plane curve

$$\begin{aligned} \dot{\mathbf{t}} &= k\mathbf{n} \\ \dot{\mathbf{n}} &= -k\mathbf{t} \end{aligned} \tag{3}$$

hold (we omit the hint about the argument s) which describe the instantaneous rotation of the moving basis.

Formulas (3) are called *Frenet's formulas for a plane curve.*

Remark 1. In an oriented plane the Frenet basis can also be defined for curves with points of rectification, taking as $\mathbf{n}(s)$ a vector forming together with $\mathbf{t}(s)$ a positively oriented basis for the plane. Then in formulas (3) a relative curvature k_{rel} appears instead of the curvature k.

For curves in a three-dimensional space (referred to rectangular coordinates x, y, z) the formula for curvature is as follows:

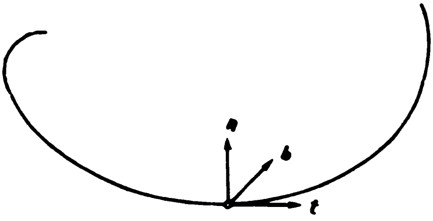
$$k = \sqrt{\ddot{x}^2 + \ddot{y}^2 + \ddot{z}^2}.$$

Curve (1) for $n = 3$, as for $n = 2$, is called a *generic curve* if $k(s) \neq 0$ for all s. For such a curve a unit vector

$$\mathbf{n}(s) = \frac{\dot{\mathbf{t}}(s)}{k(s)}$$

is defined which is called the *vector of principal normal* to the curve.

But now we can (assuming the space to be oriented) introduce a third vector, $\mathbf{b}(s)$, which forms together with $\mathbf{t}(s)$ and $\mathbf{n}(s)$ a positively oriented orthonormal basis $\mathbf{t}(s)$, $\mathbf{n}(s)$, $\mathbf{b}(s)$ (i.e. such that $\mathbf{b}(s) = \mathbf{t}(s) \times \mathbf{n}(s)$).



The Frenet basis for a space curve

This vector is called a *binormal vector*, and the basis $\mathbf{t}(s)$, $\mathbf{n}(s)$, $\mathbf{b}(s)$ *is Frenet's moving basis* for a given generic curve.

By construction (to simplify the formulas we omit the argument s)

$$\dot{\mathbf{t}} = k\mathbf{n}.$$

Moreover, since $\mathbf{b} = \mathbf{t} \times \mathbf{n}$, we have

$$\dot{\mathbf{b}} = \dot{\mathbf{t}} \times \mathbf{n} + \mathbf{t} \times \dot{\mathbf{n}} = \mathbf{t} \times \dot{\mathbf{n}}$$

from which it follows that $\dot{\mathbf{b}}\mathbf{t} = 0$. Since by Lemma 2, in Lecture 1, $\dot{\mathbf{b}}\mathbf{b} = 0$, this proves that $\dot{\mathbf{b}}$ is collinear with $\mathbf{n}$, i.e. there is a number $\varkappa = \varkappa(s)$ such that

$$\dot{\mathbf{b}} = -\varkappa\mathbf{n}.$$

The number $\varkappa(s)$ is called the *torsion* of a given curve at $\mathbf{r}(s)$. It is the rate of turn of a binormal vector.

On differentiating $\mathbf{n}\mathbf{t} = 0$ and $\mathbf{n}\mathbf{b} = 0$ we immediately get $\dot{\mathbf{n}}\mathbf{t} = -\mathbf{n}\dot{\mathbf{t}} = -k$ and $\dot{\mathbf{n}}\mathbf{b} = -\mathbf{n}\dot{\mathbf{b}} = \varkappa$. Since, in addition, $\dot{\mathbf{n}}\mathbf{n} = 0$ (Lemma 2 in Lecture 1), this proves that

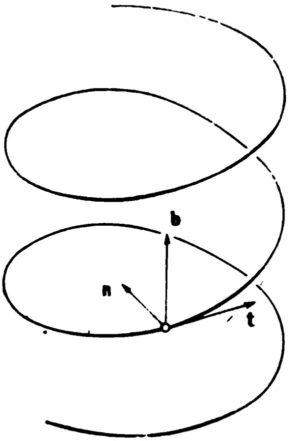
$$\dot{\mathbf{n}} = -k\mathbf{t} + \varkappa\mathbf{b}.$$

Thus, *for any generic curve we have*

$$\begin{aligned} \dot{\mathbf{t}} &= k\mathbf{n}, \\ \dot{\mathbf{n}} &= -k\mathbf{t} + \varkappa\mathbf{b}, \qquad (4) \\ \dot{\mathbf{b}} &= -\varkappa\mathbf{n}. \end{aligned}$$

These formulas are called *Frenet's formulas for a space curve.*

Example 3. If a curve $\mathbf{r} = \mathbf{r}(s)$ is in a plane Π, then the vectors $\dot{\mathbf{r}}(s)$ and $\ddot{\mathbf{r}}(s)$ are parallel to that plane (for this is the case for increments $\mathbf{r}(s + \Delta s) - \mathbf{r}(s)$ and $\dot{\mathbf{r}}(s + \Delta s) - \dot{\mathbf{r}}(s)$ of the vectors $\mathbf{r}(s)$ and $\dot{\mathbf{r}}(s)$). Therefore $\mathbf{t}(s)$, $\mathbf{n}(s) \parallel \Pi$ and hence $\mathbf{b}(s) \perp \Pi$. This proves that $\mathbf{b}(s) =$



A circular helix

const and therefore $\varkappa(s) = 0$ for all s. Conversely, let $\varkappa(s) = 0$ for all s and hence $\mathbf{b}(s) = \mathbf{b}_0 = \text{const}$. Then $(\mathbf{r}(s)\,\mathbf{b}_0)^{\cdot} = \dot{\mathbf{r}}(s)\,\mathbf{b}_0 = \mathbf{t}(s)\,\mathbf{b}_0 = 0$ for all s and therefore $\mathbf{r}(s)\,\mathbf{b}_0 = \text{const}$. This means that the curve $\mathbf{r} = \mathbf{r}(s)$ is in the plane $\mathbf{r}\mathbf{b}_0 = \text{const}$. Thus a *curve in space is a plane curve if and only if its torsion is identically zero.*

Example 4. A *circular helix* is the path described by a point moving with constant velocity around a generator of a right circular cylinder rotating uniformly about its axis. The parametric equations of the circular helix are of the form

$$x = a\cdot\cos t, \quad y = a \sin t, \quad z = bt.$$

Since $x' = -a \sin t$, $y' = a \cos t$, $z' = b$ we have

$$s' = \sqrt{(x')^2 + (y')^2 + (z')^2} = \sqrt{a^2 + b^2}$$

and hence $s = ct$, where $c = \sqrt{a^2 + b^2}$. Therefore

$$x = a \cos \frac{s}{c}, \quad y = a \sin \frac{s}{c}, \quad z = \frac{b}{c} s.$$

But then

$$\dot{x} = -\frac{a}{c} \sin \frac{s}{c}, \quad \dot{y} = \frac{a}{c} \cos \frac{s}{c}, \quad \dot{z} = \frac{b}{c},$$

$$\ddot{x} = -\frac{a}{c^2} \cos \frac{s}{c}, \quad \ddot{y} = -\frac{a}{c^2} \sin \frac{s}{c}, \quad \ddot{z} = 0$$

and hence

$$k = \sqrt{\ddot{x}^2 + \ddot{y}^2 + \ddot{z}^2} = \frac{a}{c^2} = \text{const}.$$

Besides,

$$\mathbf{t} = \left(-\frac{a}{c} \sin \frac{s}{c}, \frac{a}{c} \cos \frac{s}{c}, \frac{b}{c}\right),$$

$$\mathbf{n} = \left(-\cos \frac{s}{c}, -\sin \frac{s}{c}, 0\right)$$

and

$$\mathbf{b} = \mathbf{t} \times \mathbf{n} = \left(\frac{b}{c} \sin \frac{s}{c}, -\frac{b}{c} \cos \frac{s}{c}, \frac{a}{c}\right).$$

Therefore

$$\dot{\mathbf{b}} = \left(\frac{b}{c^2} \cos \frac{s}{c}, \frac{b}{c^2} \sin \frac{s}{c}, 0\right) = -\frac{b}{c^2}\,\mathbf{n}$$

and consequently

$$\varkappa = \frac{b}{c^2} = \text{const}.$$

Thus the *curvature and torsion of a circular helix are constant.*

According to the general Theorem 1 to be proved below, and conversely, *every curve whose curvature and torsion are constant is a circular helix* (or its arc).

Remark 2. Note the difference in the treatment of the concept of a generic curve in the plane and in space. To achieve unity, for curves in the plane we must consider relative rather than absolute curvature. Cf. Remark 1.

To investigate the behaviour of an arbitrary space curve $\mathbf{r} = \mathbf{r}(s)$ near one of its points we choose the origin O at that point, take the moving basis $\mathbf{t}_0$, $\mathbf{n}_0$, $\mathbf{b}_0$ at O as the coordinate basis $\mathbf{i}$; $\mathbf{j}$, $\mathbf{k}$ and count off the natural parameter s from 0. Then

$$\mathbf{r}(0) = 0, \quad \dot{\mathbf{r}}(0) = \mathbf{t}_0 = \mathbf{i}, \quad \ddot{\mathbf{r}}(0) = k_0\mathbf{n}_0 = k_0\mathbf{j},$$
$$\dddot{\mathbf{r}}(0) = \dot{k}_0\mathbf{n}_0 + k_0\dot{\mathbf{n}}_0 = -k^2\mathbf{i} + \dot{k}_0\mathbf{j} + k_0\varkappa_0\mathbf{k},$$

where k_0, $\dot{k}_0$ and $\varkappa_0$ are the values of the functions k, $\dot{k}$ and $\varkappa$ for $s = 0$. Hence, using the Taylor formula,

$$\mathbf{r}(s) = \mathbf{r}(0) + s\dot{\mathbf{r}}(0) + \frac{s^2}{2}\ddot{\mathbf{r}}(0) + \frac{s^3}{6}\dddot{\mathbf{r}}(0) + \dots$$
$$= (s + \dots)\,\mathbf{i} + \left(\frac{k_0}{2}s + \dots\right)\mathbf{j} + \left(\frac{k_0\varkappa_0}{6}s^3 + \dots\right)\mathbf{k}.$$

This means that near O our curve is given by the parametric equations

$$x = s + \dots$$
$$y = \frac{k_0}{2}s^2 + \dots$$
$$z = \frac{k_0\varkappa_0}{6}s^3 + \dots$$

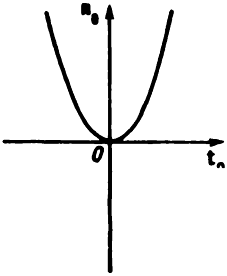
If $k_0 \neq 0$, $\varkappa_0 \neq 0$, then the projection of the curve onto the plane $O\mathbf{ij} = O\mathbf{t}_0\mathbf{n}_0$ (the plane is called the *osculating* *plane* of a space curve at O) approximately coincides with the parabola

$$x = s, \quad y = \frac{k_0}{2} s^2,$$

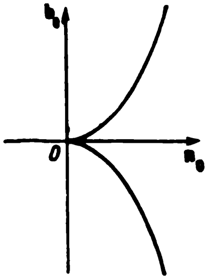
its projection onto the plane $O\mathbf{jk} = O\mathbf{n}_0\mathbf{b}_0$ (called the *normal plane* to a space curve at O) coincides with the semicubical parabola

$$y = \frac{k_0}{2} s^2, \quad z = \frac{k_0\varkappa_0}{6} s^3,$$

and, finally, its projection onto the plane $O\mathbf{ik} = O\mathbf{t}_0\mathbf{b}_0$



Projection onto an osculating plane



Projection onto a normal plane

(called a *rectifying plane* of a space curve at O) coincides with the cubical parabola

$$x = s, \quad z = \frac{k_0\varkappa_0}{2} s^3.$$

This gives a fairly clear idea of how a space curve is constructed near any of its points (at which curvature and torsion are nonzero).

Let us now consider the general case of a Euclidean space of an arbitrary dimension $n \geqslant 2$.

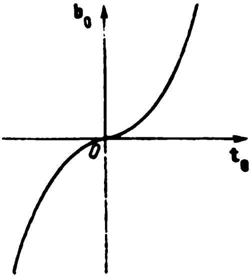
A curve $\mathbf{r} = \mathbf{r}(s)$ (referred to the natural parameter) in an n-dimensional oriented Euclidean space is called *a generic curve* if for any s the vectors

(5) $\dot{\mathbf{r}}(s), \dots, \overset{(n-1)}{\mathbf{r}}(s)$

are linearly independent.

By applying to vectors (5) the Gram-Schmidt orthogonalization process we obtain an orthonormal family of vectors $\mathbf{t}_1(s), \dots, \mathbf{t}_{n-1}(s)$. Let $\mathbf{t}_n(s)$ be a vector (uniformly defined) extending that family to a positively oriented orthonormal basis

(6) $\mathbf{t}_1(s), \dots, \mathbf{t}_{n-1}(s), \quad \mathbf{t}_n(s)$.



Projection onto a rectifying plane

Definition 2. Basis (6) is called *Frenet's moving basis* of a generic curve at a point $\mathbf{r}(s)$.

Let

$$\dot{\mathbf{t}}_i = \sum_{j=1}^{n} \alpha_{ij}\mathbf{t}_j, \quad i = 1, \dots, n$$

(we omit the argument s to simplify the formulas). Since by construction the vector $\mathbf{t}_i$, $i = 1, \dots, n-1$, is expressed linearly in terms of the vectors $\dot{\mathbf{r}}, \dots, \overset{(i)}{\mathbf{r}}$, the vector $\dot{\mathbf{t}}_i$ is expressed linearly in terms of the vectors $\dot{\mathbf{r}}, \dots, \overset{(i+1)}{\mathbf{r}}$. Since the latter vectors can be expressed linearly in terms of $\mathbf{t}_1, \dots, \mathbf{t}_{i+1}$, this proves that $\alpha_{ij} = 0$ provided that $j > i + 1$.

On the other hand, since $\mathbf{t}_i\mathbf{t}_j = \delta_{ij}$, we have $\dot{\mathbf{t}}_i\mathbf{t}_j + \mathbf{t}_i\dot{\mathbf{t}}_j = 0$, i.e.

$$\alpha_{ij} + \alpha_{ji} = 0.$$

Therefore $\alpha_{ii} = 0$ and $\alpha_{ij} = 0$ for $j < i - 1$.

Thus only the coefficients $\alpha_{i,i+1} = -\alpha_{i+1,i}$ can be nonzero. Setting

$$k_1 = \alpha_{12}, \quad k_2 = \alpha_{23}, \dots, k_{n-1} = \alpha_{n-1,n}$$

we therefore see that the following formulas hold

$$
\begin{aligned}
&\dot{\mathbf{t}}_1 = k_1\mathbf{t}_2,\\
&\dot{\mathbf{t}}_2 = -k_1\mathbf{t}_1 + k_2\mathbf{t}_3,\\
(7)\qquad &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\\
&\dot{\mathbf{t}}_{n-1} = -k_{n-2}\mathbf{t}_{n-2} + k_{n-1}\mathbf{t}_n,\\
&\dot{\mathbf{t}}_n = -k_{n-1}\mathbf{t}_{n-1}.
\end{aligned}
$$

These formulas are called *Frenet's formulas for a curve in an n-dimensional space*.

The functions $k_1 = k_1(s), \ldots, k_{n-1} = k_{n-1}(s)$ are called the *curvatures* of a curve. They are defined only for a generic curve.

In the formulas

$$
(8)\qquad \mathbf{t}_i = \beta_{i1}\dot{\mathbf{r}} + \ldots + \beta_{ii}\overset{(i)}{\mathbf{r}}_i, \quad i = 1, \ldots, n-1,
$$

resulting from applying the Gram-Schmidt orthogonalization process, the coefficients β_{ii} are positive. Therefore in the reverse formulas

$$
(9)\qquad \overset{(i)}{\mathbf{r}} = \gamma_{i1}\mathbf{t}_1 + \ldots + \gamma_{ii}\mathbf{t}_i
$$

the coefficients $\gamma_{ii} = \beta_{ii}^{-1}$ are also positive. Differentiating formulas (8) we get

$$
\dot{\mathbf{t}}_i = \dot{\beta}_{i1}\dot{\mathbf{r}} + (\dot{\beta}_{i2} + \beta_{i1})\ddot{\mathbf{r}} + \ldots + (\dot{\beta}_{ii} + \beta_{i,i-1})\overset{(i)}{\mathbf{r}} + \beta_{ii}\overset{(i+1)}{\mathbf{r}},
$$
$$
i = 1, \ldots, n-1.
$$

On replacing here (for $i < n-1$) the vectors $\dot{\mathbf{r}}, \ldots, \overset{(i+1)}{\mathbf{r}}$ by expressions (9) we must get formulas (7). This shows that

$$
k_i = \beta_{ii}\gamma_{i+1,i+1}, \quad i = 1, \ldots, n-2.
$$

It follows, in particular, that *for any curve of the general type the curvatures*

$$
k_1, \ldots, k_{n-2}
$$

are positive. The curvature k_{n-1} (the analogue of torsion), on the other hand, may have any sign.

Now we show that any $n - 1$ functions

$$k_1(s) > 0, \quad \ldots, \quad k_{n-2}(s) > 0, \quad k_{n-1}(s) \tag{10}$$

(given on some interval (a, b) of the axis $\mathbb{R}$) may serve as curvatures for a curve (regular but generally not simple) and that these curvatures uniquely (up to congruence) define the curve.

Let for definiteness $a < 0 < b$.

Theorem 1. *Let $n - 1$ smooth functions (10), all positive except possibly the last, be given on an arbitrary interval (a, b). Then for any initial point $O \in \mathcal{A}$ and for any positively oriented orthonormal basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$ there exists one and only one generic curve $\mathbf{r} = \mathbf{r}(s)$, $a < s < b$ of the general type having the following two properties:*

1° *the given functions (10) are the curvatures of the curve,*
2° *for $s = 0$ we have*

$$\mathbf{r}(0) = 0, \quad \mathbf{t}_1 = (0) = \mathbf{i}_1, \quad \ldots, \quad \mathbf{t}_n(0) = \mathbf{i}_n.$$

Proof. We carry out the proof in four stages.

Stage 1. At this stage we use the **unique existence theorem** for solutions (UES) of linear ordinary differential equations.

Theorem (UES). *Let m^2 smooth functions $A_{ij}(s)$, $i, j = 1, \ldots, m$ be given on an interval (a, b) and let $x_1^{(0)}, \ldots, x_m^{(0)}$ be arbitrary numbers. Then there is one and only one family of smooth functions $x_1(s), \ldots, x_m(s)$, $a < s < b$, having the following two properties:*

1° *identically in terms of s, $a < s < b$, the relations*

$$\begin{aligned} \dot{x}_1 &= A_{11}x_1 + \ldots + A_{1m}x_m, \\ &\ldots\ldots\ldots\ldots\ldots\ldots, \\ \dot{x}_m &= A_{m1}x_1 + \ldots + A_{mm}x_m \end{aligned} \tag{11}$$

hold,

2° *for $s = 0$ we have*

$$x_1(0) = x_0^{(0)}, \ \ldots, \ x_m(0) = x_m^{(0)}. \ \square$$

We shall apply this theorem to relations (7) which for the given functions $k_1, \ldots, k_{n-1}$ are equations of the form (11) for $m = n^2$ coordinate vectors $\mathbf{t}_1, \ldots, \mathbf{t}_n$. Thus, according to the UES theorem, there is one and only one family of vector-valued functions $\mathbf{t}_1(s), \ldots, \mathbf{t}_{n-1}(s)$, $a < s < b$, on (a, b) such that

1° for any s relations (7) are satisfied,

2° for $s = 0$ equations

$$\mathbf{t}_1(0) = \mathbf{i}_1, \ldots, \mathbf{t}_n(0) = \mathbf{i}_n \tag{12}$$

hold.

Stage 2. We consider scalar products $\mathbf{t}_i\mathbf{t}_j$, $i, j = 1, \ldots, n$. According to relations (7), for these products we have

$$\begin{aligned}(\mathbf{t}_i\mathbf{t}_j)^{\cdot} &= \dot{\mathbf{t}}_i\mathbf{t}_j + \mathbf{t}_i\dot{\mathbf{t}}_j\\ &= (-k_{i-1}\mathbf{t}_{i-1} + k_i\mathbf{t}_{i+1})\,\mathbf{t}_j\\ &\quad + \mathbf{t}_i\,(-k_{j-1}\mathbf{t}_{j-1} + k_j\mathbf{t}_{j+1})\end{aligned}$$

(we assume that $\mathbf{t}_0 = 0$ and $\mathbf{t}_{n+1} = 0$), i.e. the equations

$$\begin{aligned}(\mathbf{t}_i\mathbf{t}_j)^{\cdot} = &-k_{i-1}(\mathbf{t}_{i-1}\mathbf{t}_j) + k_i(\mathbf{t}_{i+1}\mathbf{t}_j)\\ &-k_{j-1}(\mathbf{t}_i\mathbf{t}_{j-1}) + k_j(\mathbf{t}_i\mathbf{t}_{j+1})\end{aligned} \tag{13}$$

which may be regarded as equations of the form (11) for $m = \frac{n(n+1)}{2}$ functions $\mathbf{t}_i\mathbf{t}_j$. By the UES theorem therefore there is one and only one set of these functions having the property that for $s = 0$ they are equal to $\delta_{ij} = \mathbf{i}_i\mathbf{i}_j$ (i.e. to zero if $i \neq j$ and to unity if $i = j$).

On the other hand, a direct verification shows that the functions $\mathbf{t}_i\mathbf{t}_j$ identically equal to δ_{ij} satisfy equations (13). (Indeed, when $i \neq j-1, j+1$ all the terms of the sum $-k_{i-1}\delta_{i-1,j} + k_i\delta_{i+1,j} - k_{j-1}\delta_{i,j-1} + k_j\delta_{i,j+1}$ are zero and when $i = j-1, j+1$ the sum has only two terms which are nonzero but cancel in pairs.) Hence, for all s there are, by virtue of the UES theorem, equations $\mathbf{t}_i\mathbf{t}_j = \delta_{ij}$, $i, j = 1, \ldots, n$, implying that for any s, $a < s < b$, the vectors $\mathbf{t}_1, \ldots, \mathbf{t}_n$ form an orthonormal basis.

Since for $s = 0$ that basis coincides with a positively oriented basis $\mathbf{i}_1, \ldots, \mathbf{i}_n$, the basis $\mathbf{t}_1, \ldots, \mathbf{t}_n$ is positively oriented for any s, $a < s < b$, too.

Stage 3. We compose consecutive derivatives for the vector $\mathbf{t}_1$

$$(14) \qquad \mathbf{t}_1,\ \dot{\mathbf{t}}_1,\ \ddot{\mathbf{t}}_1,\ \ldots,\ \overset{(n-1)}{\mathbf{t}_1}$$

and apply the Gram-Schmidt orthogonalization process to them. Since $\mathbf{t}_1$ is a unit vector, we need not do anything in the first step of the process. Since by Lemma 2 in Lecture 1 $\dot{\mathbf{t}}_1$ is orthogonal to $\mathbf{t}_1$, in the second step we must only normalize it. However, since we have proved $\mathbf{t}_2$ to be a unit vector and under the hypothesis $k_1 > 0$, according to the first of the relations (7) $|\dot{\mathbf{t}}_1| = k_1$. In the second step therefore we obtain the vector

$$\mathbf{t}_2 = \frac{\dot{\mathbf{t}}_1}{k_1}.$$

In the third step we should consider the vector

$$\ddot{\mathbf{t}}_1 = (k_1\mathbf{t}_2)^{\cdot} = \dot{k}_1\mathbf{t}_2 + k_1\dot{\mathbf{t}}_2 = -k_1^2\mathbf{t}_1 + \dot{k}_1\mathbf{t}_2 + k_1k_2\mathbf{t}_3,$$

subtract from it the linear combination of vectors $\mathbf{t}_1$ and $\mathbf{t}_2$ so as to obtain a vector orthogonal to those vectors and then normalize the vector. But since according to what has been proved the vectors $\mathbf{t}_1$, $\mathbf{t}_2$, $\mathbf{t}_3$ form an orthonormal family and under the hypothesis $k_1k_2 > 0$, the procedure will obviously yield the vector $\mathbf{t}_3$.

It is clear that this reasoning is of a general character so that at each step of the orthogonalization process we obtain the corresponding vector $\mathbf{t}_i$, $i = 1, \ldots, n-1$. This proves that the family of vectors $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{n-1}$ is uniquely characterized as an orthonormal family of vectors obtained from family (14) by applying Gram-Schmidt orthogonalization.

Stage 4. Let

$$(15) \qquad \mathbf{r}(s) = \int_0^s \mathbf{t}_1(s)\,ds, \qquad a < s < b.$$

Then $\dot{\mathbf{r}}(0) = 0$ and $\dot{\mathbf{r}}(s) = \mathbf{t}_1(s)$, i.e. for $s = 0$ the curve $\mathbf{r} = \mathbf{r}(s)$, $a < s < b$ passes through the point O and for any s the vector $\mathbf{t}_1(s)$ is tangent to it. But for every curve the first $n - 1$ vectors of the moving basis are vectors obtained from the first $n - 1$ derivatives of the tangent vector by the Gram-Schmidt orthogonalization process. According to the foregoing therefore those vectors coincide with $\mathbf{t}_1, \ldots, \mathbf{t}_{n-1}$.

As to the last vector of the moving basis, it is uniquely characterized as a unit vector which together with the first $n - 1$ vectors forms a positively oriented basis. Since the basis $\mathbf{t}_1, \ldots, \mathbf{t}_n$, as we have seen, is positively oriented, the vector $\mathbf{t}_n$ should be the required vector.

Thus, we have proved that for any s the vectors $\mathbf{t}_1(s), \ldots, \mathbf{t}_n(s)$ form the moving basis of the curve $\mathbf{r} = \mathbf{r}(s)$. Since for these vectors the Frenet formulas (7) are valid, the functions $k_i(s)$, $i = 1, \ldots, n-1$ appearing in the formulas must be the curvatures of $\mathbf{r} = \mathbf{r}(s)$.

This completes the proof of the existence of a curve $\mathbf{r} = \mathbf{r}(s)$ having Properties 1° and 2°.

The uniqueness of the curve follows from the fact that according to the UES theorem, the moving basis $\mathbf{t}_1(s), \ldots, \mathbf{t}_n(s)$ is uniquely defined by equations (7) and the initial conditions (12) and the radius vector $\mathbf{r}(s)$ is uniquely defined (by formula (15)) by the relation $\dot{\mathbf{r}}(s) = \mathbf{t}_1(s)$ and the initial condition $\mathbf{r}(0) = 0$. □

Lecture 3

Elementary surfaces and their parametrizations · Examples of surfaces · Tangent plane and tangent subspace · Smooth mappings of surfaces and their differentials · Diffeomorphisms of surfaces · The first quadratic form of a surface · Isometries · Beltrami's first differential parameter · Examples of computation of first quadratic forms · Developable surfaces

---

An intuitive idea of a surface is explicated by analogy with that of a line, but the explication is more difficult that that for a line. We shall restrict our discussion to the analogue of an idea of an open simple regular arc (though considering more general surfaces when discussing particular examples).

To introduce this analogue we begin with an arbitrary continuous mapping of the form

$$\gamma\colon\ U \to \mathscr{A}, \tag{1}$$

where $\mathscr{A}$ is some Euclidean (or only affine) space of dimension $n \geqslant 3$ and U is a *convex* (i.e. containing every straight segment whose ends are in the set) open subset of a Euclidean plane $\mathbb{R}^2$ (a two-dimensional analogue of the interval $I = (a, b)$). Once the reference point O has been chosen in $\mathscr{A}$, mapping (1) is given by a continuous vector function

$$\mathbf{r} = \mathbf{r}(u, v), \quad (u, v) \in U \tag{2}$$

which assumes values in the associated vector space $\mathscr{V}$ and when, in addition, a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is chosen in $\mathscr{V}$, (1) is given by n continuous numerical functions

$$x^1 = x^1(u, v), \ \ldots, \ x^n = x^n(u, v), \tag{3}$$

the coordinates in the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of the vector $\mathbf{r}(u, v)$.

Mapping (or map) (1) is said to be a *C^r-mapping*, where r is some natural number or a symbol ∞, if every function (3) or equivalently vector function (2) has continuous partial derivatives of all orders $\leqslant r$ (recall that when $r = \infty$ this implies the existence of partial derivatives of all orders). In what follows we shall assume the number r to be once and for ever fixed and sufficiently large. [In this lecture any number $r \geqslant 1$ will suit, but in, say, Lecture 5 we shall have to require that $r \geqslant 3$.]

In particular, for the smooth mapping (1) partial derivatives

$$\mathbf{r}_u = \frac{\partial \mathbf{r}(u, v)}{\partial u}, \quad \mathbf{r}_v = \frac{\partial \mathbf{r}(u, v)}{\partial v} \tag{4}$$

of the vector function (2) are defined. Smooth mapping (1) is said to be *regular* if at every point $(u, v) \in U$ partial derivatives (4) are linearly independent.

Definition 1. Mapping (1) is said to be a *parametrization* if it is:

1° smooth,

2° regular, and

3° moneomorphic (injective and has the property that if a sequence of points $\gamma(u_n, v_n)$, $(u_n, v_n) \in U$, of $\mathcal{A}$ converges to a point of the form $\gamma(a, b)$, where $(a, b) \in U$, then the sequence of points $(u_n, v_n) \in U$ also converges (to the point (a, b), by virtue of continuity).

Problem 1. Prove that if mapping (1) is smooth and regular, then for any point $(u_0, v_0) \in U$ there is its neighbourhood $V \subset U$ on which this mapping is moneomorphic (is a parametrization).

Definition 2. A subset $\mathscr{X}$ of a space $\mathcal{A}$ is said to be an *elementary surface* if there is a parametrization $\gamma: U \to \mathcal{A}$ (called in this case a *parametrization of the surface $\mathscr{X}$*) such that $\gamma(U) = \mathscr{X}$. It is also said that $\mathscr{X}$ is the *support* of a parametrization γ.

Elementary surfaces are the two-dimensional analogues of simple regular arcs (and parametrizations are the analogues of simple regular curves).

Remark 1. In another terminological scheme, which we shall also use now and again, sometimes without stating this explicitly, parametrizations (1) themselves are called the elementary surfaces. For lines we have avoided such homonyms by distinguishing between curves and lines. Unfortunately, there is no such pair of generally accepted terms for the two-dimensional case.

Since, as a rule, we consider only elementary surfaces in what follows we shall call elementary surfaces simply surfaces.

Since the parametrization $\gamma\colon U \to \mathcal{A}$ of an arbitrary elementary surface $\mathcal{X}$ is an injective mapping, for any point $p \in \mathcal{X}$ there are unique numbers u and v having the property that $(u, v) \in U$ and $\gamma(u, v) = p$. These numbers are called the *coordinates* of p in the given parametrization. Traditionally additional terms are used for these coordinates, they are called *curvilinear* or *local*, although no other coordinates on the surface are usually considered, and therefore these terms are not necessary.

Speaking loosely, numbers u and v are often called the coordinates *on a surface* (1); this is a manifestation of a general tendency to confuse in usage surfaces and their parametrizations.

Every curve in U with parametric equations

$$u = u(t), \quad v = v(t), \quad t \in I, \tag{5}$$

is mapped by parametrization (1) of a surface $\mathcal{X}$ to a curve

$$\mathbf{r} = \mathbf{r}(u(t), v(t)), \quad t \in I, \tag{6}$$

of a space $\mathcal{A}$. Curve (6) is said to *lie* on $\mathcal{X}$ and equations (5) to be its *parametric equations in coordinates u and v.*

In particular, curves $u = \text{const}$ and $v = \text{const}$ (which are the images of coordinate lines in U) are called the *coordinate lines on the surface* $\mathcal{X}$ and their totality is the *coordinate network*.

For any open subsets U, $U^* \subset \mathbb{R}^2$ every mapping

$$\varphi\colon U^* \to U \tag{7}$$

is given by a pair of functions

$$u = u(u^*, v^*), \quad v = v(u^*, v^*), \quad (u^*, v^*) \in U^* \tag{8}$$

having the property that for any point $(u^*, v^*) \in U^*$ the point $(u, v) = (u(u^*, v^*), v(u^*, v^*))$ is in U. Mapping (7) is said to be a C^r*-mapping*, where r is a natural number or a symbol ∞, if for $r \neq \infty$ functions (8) have continuous partial derivatives of all orders $\leqslant r$. Smooth mapping (7) is said to be a diffeomorphism if it is bijective and the inverse mapping

$$\varphi^{-1}\colon U \to U^* \tag{8}$$

is also smooth.

Two parametrizations

$$\gamma\colon U \to \mathcal{A} \quad \text{and} \quad \gamma^*\colon U^* \to \mathcal{A} \tag{9}$$

are said to be *equivalent* if there is a diffeomorphism (7) such that

$$\gamma^* = \gamma \circ \varphi. \tag{10}$$

Since giving a parametrization of a surface $\mathscr{X}$ is equivalent to giving curvilinear coordinates on $\mathscr{X}$, diffeomorphism (7) is also said to perform a *change of coordinates* on $\mathscr{X}$ (or to specify a *transition* from coordinates u, v to coordinates u^*, v^*).

Problem 2. Prove that relation (10) is an equivalence in the general algebraic sense (is reflexive, symmetric and transitive) and hence we should speak about classes of equivalent parametrizations.

It is clear that *equivalent parametrizations have the same support*. Conversely, we can easily show that *parametrizations* (9) *having the same support are equivalent*. This means that elementary surfaces are bijectively associated with equivalence classes of their parametrizations and therefore can be identified with them.

Problem 3. Prove the last statement. (Notice that for this to be true all the three conditions 1° to 3° of Definition 1 are essential. Cf. the proof of Proposition 1 of Lecture 1.)

In Lecture 15 we prove the general proposition of which this statement (as well as Proposition 1 of Lecture 1) is a special case.

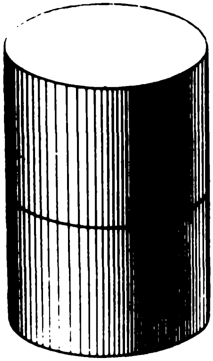
Examples of surfaces.

For clarity we restrict ourselves to surfaces in a three-dimensional Euclidean space. The coordinates x, y, z will be assumed to be rectangular.

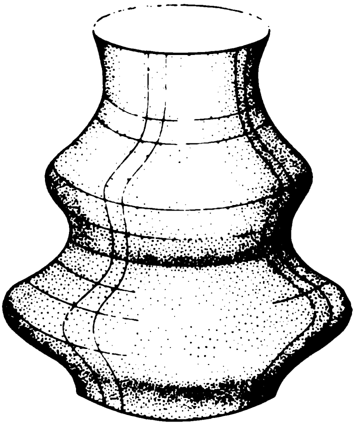
Example 1. The equations

$$x = R\cos u, \quad y = R\sin u, \quad z = v \tag{11}$$

give a right circular cylinder in a three-dimensional Euclidean space. This is not an elementary surface in the sense of Definition 2, since for $-\infty < u < +\infty$ each point of the cylinder is covered an infinite (countable) number of times by the points of the plane $\mathbb{R}^2$. To obtain an elementary surface the cylinder should be cut through along its element, i.e. in (11) the parameter u should satisfy the inequalities $0 < u < 2\pi$. The entire cylinder, however, is covered by two such cut cylinders.



A circular cylinder



A surface of revolution

The coordinate network on cylinder (11) consists of vertical straight lines $u = \text{const}$ and horizontal circles $v = \text{const}$.

Example 2. Let $x = x(v)$, $z = z(v)$ be a simple regular curve in the plane Oxz not intersecting the Oz-axis. The surface with parametrization

$$(12) \quad x = x(v)\cos u, \quad y = x(v)\sin u, \quad z = z(v)$$

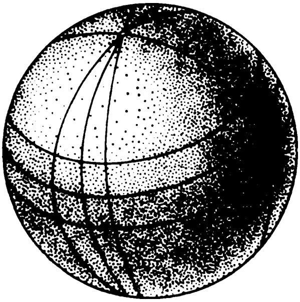
is called a *surface of revolution* and the curve $x = x(v)$, $z = z(v)$ is its *profile*. Graphically, surface (12) is obtained by rotating its profile about the Oz-axis.

The regularity of parametrization (12), i.e. linear independence of vectors

$$\mathbf{r}_u = (-x(v)\sin u,\ x(v)\cos u,\ 0),$$
$$\mathbf{r}_v = (x'(v)\cos u,\ x'(v)\sin u,\ z'(v))$$

is ensured by the regularity of the profile (i.e. by the condition $x'(v)^2 + z'(v)^2 = 1$) and by the fact that the profile does not intersect the Oz-rotation axis (i.e. by $x(v) \neq 0$).

The coordinate network on surface (12) consists of plane curves which are rotations of the profile about the Oz-axis (they are called *meridians*) and circles perpendicular to them (*parallels*). To make surface (12) elementary, it should be cut through along a meridian.



A sphere

A cylinder is a surface of revolution whose profile is a straight line $x = R$, $z = v$.

A sphere

$$x = R\cos v\cos u,\quad y = R\cos v\sin u,\quad z = R\sin v$$

of radius R with centre at O is a surface of revolution of a circle with profile $x = R\cos v$, $z = R\sin v$. The coordinates u and v on the sphere are the well-known "geographical coordinates", longitude and latitude, and the coordinate curves are geographical meridians and parallels.

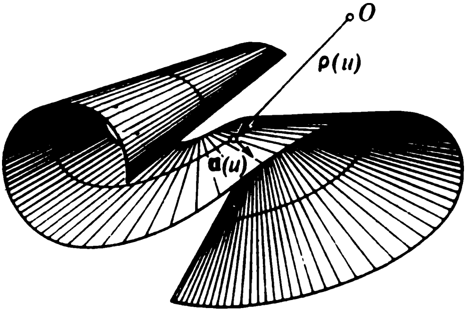
Strictly speaking, the profile of a sphere is only its semicircle $-\pi < v < +\pi$ (which excludes the poles).

To obtain an elementary surface it is necessary to exclude one meridian (the "line of change of dates").

Example 3. A surface $\mathbf{r} = \mathbf{r}(u, v)$ is said to be a *ruled surface* if

$$\mathbf{r}(u, v) = \boldsymbol{\rho}(u) + v\mathbf{a}(u), \tag{13}$$

where $\boldsymbol{\rho}(u)$ and $\mathbf{a}(u)$ are vector-valued functions having the property (ensuring regularity) that the vectors



A ruled surface

$\boldsymbol{\rho}'(u) + v\mathbf{a}'(u)$ and $\mathbf{a}(u)$ are linearly independent for all u and v under consideration (so that, in particular, $\mathbf{a}(u) \neq 0$ for all u). A straight line with a direction vector $\mathbf{a}(u_0)$, which passes through the point with radius vector $\boldsymbol{\rho}(u_0)$ is a coordinate curve $u = u_0 = \text{const}$. Thus, graphically, a ruled surface is generated by various positions of a moving straight line. Cf. Definition 1, Lecture I.23.

It is clear that we may assume without loss of generality the vector $\mathbf{a}(u)$ to be a unit vector:

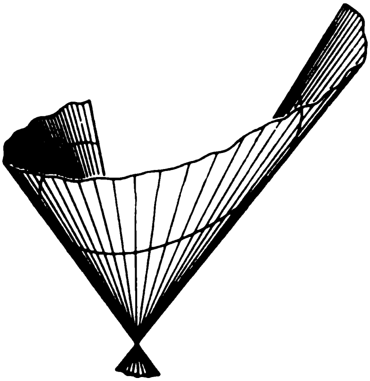
$$|\mathbf{a}(u)| = 1 \text{ for all } u.$$

If $\boldsymbol{\rho}'(u) = 0$ for all u, i.e. $\boldsymbol{\rho}(u) = \text{const}$, then, after translation of the origin, we obtain instead of (13) an equation of the form

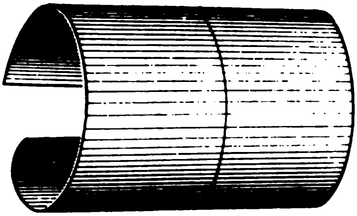
$$\mathbf{r} = v\mathbf{a}(u). \tag{14}$$

It is a cone whose directrix is a regular space curve $\mathbf{r} = \mathbf{a}(u)$. [In (14) it must be assumed that $v > 0$ or $v < 0$ (for $v = 0$ is a singular point of the cone which divides it into two nappes). But if the generators of the cone intersect its directrix at several points, additional constraints have to be introduced on v.]

If $\mathbf{a}'(u) = 0$ for all u, i.e. $\mathbf{a}(u) = \text{const}$, then surface (14) is a cylinder with directrix $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ (generally a space one).



A cone

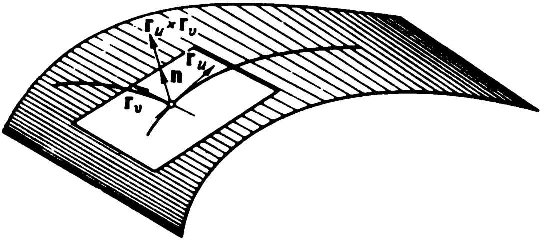


A cylinder

If the vector $\boldsymbol{\rho}'$ is not identically zero, then, going over if necessary to a smaller domain in $\mathbb{R}^2$, we may assume that $\boldsymbol{\rho}'(u) \neq 0$ for all u. Then $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ is a regular curve in space and we may assume that u is the natural parameter (arc length) on that curve. Cone (14) may also be specified by an equation of form (13) with $\boldsymbol{\rho}'(u) \neq 0$. To do this it is sufficient to put $\boldsymbol{\rho}(u) = \mathbf{a}(u)$ in (13) (if $\mathbf{a}'(u) \neq 0$).

If $\mathbf{a}(u)$ is a tangent vector $\boldsymbol{\tau}(u)$ of a curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$, then surface (13) is called a *surface of tangents*. A *surface of principal normals* and a *surface of binormals* are defined similarly.

Notice that for a surface of tangents all points of the curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ are singular points of the surface at which the regularity condition fails. (They form the so-called *edge of regression* of the surface of tangents.)



A tangent plane

Again let $\mathscr{X}$ be an arbitrary (elementary) surface in an n-dimensional affine space $\mathscr{A}$ with parametrization $\mathbf{r} = \mathbf{r}(u, v)$ and let p_0 be an arbitrary point of $\mathscr{X}$ and $\mathbf{r}_0 = \mathbf{r}(u_0, v_0)$ be its radius vector. By the regularity condition the values

$$\mathbf{r}_{u_0} = \frac{\partial \mathbf{r}}{\partial u}(u_0, v_0), \quad \mathbf{r}_{v_0} = \frac{\partial \mathbf{r}}{\partial v}(u_0, v_0)$$

of partial derivatives of the vector function $\mathbf{r} = \mathbf{r}(u, v)$ at the point (u_0, v_0) are linearly independent, i.e. the bivector $\mathbf{r}_{u_0} \wedge \mathbf{r}_{v_0}$ is nonzero. A two-dimensional plane, which has a direction bivector $\mathbf{r}_{u_0} \wedge \mathbf{r}_{v_0}$ and passes through the point p_0 is therefore defined in $\mathscr{A}$. The vector parametric equation of that plane has the form

$$\mathbf{r} = \mathbf{r}_0 + a\mathbf{r}_{u_0} + b\mathbf{r}_{v_0}, \tag{15}$$

where a and b are parameters.

Definition 3. Plane (15) is called the *tangent plane* of a surface $\mathscr{X}$ (or *to a surface* $\mathscr{X}$) at p_0. The corresponding subspace of the associated vector space $\mathscr{V}$ (consisting of vectors of the form $a\mathbf{r}_{u_0} + b\mathbf{r}_{v_0}$) is called a *tangent* *subspace* and is denoted by $\mathbf{T}_{p_0}\mathscr{X}$ and its vectors are *tangent vectors* of $\mathscr{X}$ at p_0. To stress its two-dimensional character, the tangent subspace is often called a *tangent plane*.

This terminology is justified by the fact that for any curve (5) on $\mathscr{X}$ passing (say, for $t = t_0$) through p_0 its tangent vector

$$\mathbf{r}'(t_0) = u'(t_0)\mathbf{r}_{u_0} + v'(t_0)\,\mathbf{r}_{v_0} \tag{16}$$

is in $\mathbf{T}_{p_0}\mathscr{X}$, and conversely any vector $a\mathbf{r}_u + b\mathbf{r}_v$ in $\mathbf{T}_{p_0}\mathscr{X}$ can be represented in form (16) (it suffices to consider a curve with parametric equations $u = u_0 + at$, $v = v_0 + bt$, $t \in I$, where I is an interval of a t-axis such that $(u_0 + at, v_0 + bt) \in U$ for any $t \in I$). *Tangent vectors of a surface are thus vectors tangent to the curves on that surface.*

Relation (10) which defines a change of coordinates on $\mathscr{X}$ can be written in radius vectors as a formula

$$\mathbf{r}^*(u^*, v^*) = \mathbf{r}(u(u^*, v^*), v(u^*, v^*))$$

differentiation of which gives the relations

$$\begin{aligned} \mathbf{r}^*_{u^*} &= \frac{\partial u}{\partial u^*}\mathbf{r}_u + \frac{\partial v}{\partial u^*}\mathbf{r}_v, \\ \mathbf{r}^*_{v^*} &= \frac{\partial u}{\partial v^*}\mathbf{r}_u + \frac{\partial v}{\partial v^*}\mathbf{r}_v. \end{aligned} \tag{17}$$

It follows from relations (17) that the *vectors* $\mathbf{r}^*_{u^*}$ and $\mathbf{r}^*_{v^*}$ *are linearly equivalent to* $\mathbf{r}_u$ and $\mathbf{r}_v$ and hence generate the same subspace. This proves that *for any point $p \in \mathscr{X}$ the vector space $\mathbf{T}_p\mathscr{X}$ is defined correctly* (is independent of the choice of parametrization $\mathbf{r} = \mathbf{r}(u, v)$). When changing the parametrization in $\mathbf{T}_p\mathscr{X}$ only the basis $\mathbf{r}_u, \mathbf{r}_v$ is changed, by formulas (17).

The vectors of space $\mathbf{T}_p\mathscr{X}$ are usually denoted by $d\mathbf{r}$ and their coordinates (in the basis $\mathbf{r}_u, \mathbf{r}_v$) by du and dv (these coordinates being written to the right of the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$). Thus, in the notation

$$d\mathbf{r} = \mathbf{r}_u du + \mathbf{r}_v dv \tag{18}$$

for any vector $d\mathbf{r}$ of $\mathbf{T}_p\mathscr{X}$. It follows directly from (17) that the coordinates du, dv are connected with du^*, dv^* in the basis $\mathbf{r}^*_{u^*}$, $\mathbf{r}^*_{v^*}$ by the relations

$$\begin{aligned} du &= \frac{\partial u}{\partial u^*}\,du^* + \frac{u}{\partial v^*}\,dv^*, \\ dv &= \frac{\partial v}{\partial u^*}\,du^* + \frac{\partial v}{\partial v^*}\,dv^* \end{aligned} \tag{19}$$

which formally coincide with the formulas for differentials known from calculus (which is the basic argument for notation (18)).

Now let us consider together with the surface $\mathscr{X}$ its parametrization $\gamma\colon U \to \mathscr{A}$ given by the vector function $\mathbf{r} = \mathbf{r}(u, v)$, $(u, v) \in U$, another elementary surface $\hat{\mathscr{X}}$ with parametrization $\hat\gamma\colon \hat U \to \mathscr{A}$ given by the vector function $\mathbf{r} = \hat{\mathbf{r}}(\hat u, \hat v)$, $(\hat u, \hat v) \in \hat U$.

Since γ and $\hat\gamma$ are injective mappings, every mapping f: $\mathscr{X} \to \hat{\mathscr{X}}$ uniquely defines the mapping $\bar f\colon U \to \hat U$ which sastisfies the relation

$$f \circ \gamma = \hat\gamma \circ \bar f \tag{20}$$

and uniquely defines the mapping f. Graphically relation (20) means that in the diagram

$$\begin{tikzcd} \mathscr{X} \arrow[r, "f"] & \hat{\mathscr{X}} \\ U \arrow[u, "\gamma"] \arrow[r, "\bar f"'] & \hat U \arrow[u, "\hat\gamma"'] \end{tikzcd} \tag{21}$$

going from the bottom left-hand corner to the top right-hand corner along with the two possible paths leads to the same result. [Such diagrams are called *commutative*.]

The mapping $\bar f$ is said to *represent* a mapping f in parametrizations, γ and $\hat\gamma$ and the functions

$$\hat u = \hat u(u, v), \quad \hat v = \hat v(u, v), \quad (u, v) \in U \tag{22}$$

which specify $\bar f$ are said to *give f in coordinates u, v and u^*, v^**.

A mapping f is said to be *smooth* if so is $\bar{f}$, i.e. if so are functions (22). This definition is correct (if $\bar{f}$ is smooth for one choice of parametrizations γ and $\hat{\gamma}$, then it will be so for any other).

We associate every smooth mapping $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ and any point $p \in \mathscr{X}$ with a linear mapping

$$\mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\hat{p}}\hat{\mathscr{X}}, \quad \hat{p} = f(p) \tag{23}$$

of tangent spaces carrying vector (18) of the space $\mathbf{T}_p\mathscr{X}$ into the vector

$$d\hat{\mathbf{r}} = \mathbf{r}_{\hat{u}}\, d\hat{u} + \mathbf{r}_{\hat{v}}\, d\hat{v}$$

of the space $\mathbf{T}_{\hat{p}}\hat{\mathscr{X}}$, where in full accord with the differential calculus formulas

$$\begin{aligned} d\hat{u} &= \frac{\partial \hat{u}}{\partial u}\, du + \frac{\partial \hat{u}}{\partial v}\, dv, \\ d\hat{v} &= \frac{\partial \hat{v}}{\partial u}\, du + \frac{\partial \hat{v}}{\partial v}\, dv. \end{aligned} \tag{24}$$

Mapping (23) is defined for the given parametrizations γ, $\hat{\gamma}$ of surfaces $\mathscr{X}$, $\hat{\mathscr{X}}$ and therefore the question arises as to whether it is correct, i.e. independent of the choice of these parametrizations.

Suppose, for example, we have replaced the parametrization γ by another parametrization $\gamma^*\colon U^* \to \mathscr{A}$ of $\mathscr{X}$. By definition $\gamma^* = \gamma \circ \varphi$, where $\varphi\colon U^* \to U$ is some diffeomorphism given by the functions

$$u = u(u^*, v^*), \quad v = v(u^*, v^*).$$

For the corresponding bases $\mathbf{r}_u$, $\mathbf{r}_v$ and $\mathbf{r}_{u^*}$, $\mathbf{r}_{v^*}$ of $\mathbf{T}_p\mathscr{X}$ formulas (17) hold and for the corresponding coordinates du, dv and du^*, dv^* of the tangent vectors formulas (19) do. In coordinates u^*, v^*, and $\hat{u}$, $\hat{v}$ the mapping f is obviously specified by the functions

$$\begin{aligned} \hat{u}^*(u^*, v^*) &= \hat{u}(u(u^*, v^*),\ v(u^*, v^*)), \\ \hat{v}^*(u^*, v^*) &= \hat{v}(u(u^*, v^*),\ v(u^*, v^*)) \end{aligned}$$

and hence the vector $d\mathbf{r}$ of $\mathbf{T}_p\mathscr{X}$, which has coordinates du^*, dv^* in the basis $\mathbf{r}_{u^*}$, $\mathbf{r}_{v^*}$, will carry mapping (23) constructed using the parametrizations γ^*, $\hat{\gamma}$ into the vector $d\hat{\mathbf{r}}^*$ of $\mathbf{T}_{\hat{p}}\hat{\mathscr{X}}$ which has coordinates

$$d\hat{u}^* = \frac{\partial \hat{u}^*}{\partial u^*}du^* + \frac{\partial \hat{u}^*}{\partial v^*}\,dv^*,$$

$$d\hat{v}^* = \frac{\partial \hat{v}^*}{\partial u^*}du^* + \frac{\partial \hat{v}^*}{\partial v^*}\,dv^*$$

in the basis $\hat{\mathbf{r}}_{\hat{u}}$, $\hat{\mathbf{r}}_{\hat{v}}$. On the other hand, on substituting in formulas (24) for the coordinates $d\hat{u}$, $d\hat{v}$ of vector $d\hat{\mathbf{r}}$ expressions (19) for the coordinates du, dv in terms of du^*, dv^* and taking into account the fact that according to the rules for differentiating composite functions known from calculus the equations

$$\frac{\partial \hat{u}}{\partial u}\,\frac{\partial u}{\partial u^*} + \frac{\partial \hat{u}}{\partial v}\,\frac{\partial v}{\partial u^*} = \frac{\partial \hat{u}^*}{\partial u^*}, \qquad \frac{\partial \hat{u}}{\partial u}\,\frac{\partial u}{\partial v^*} + \frac{\partial \hat{u}}{\partial v}\,\frac{\partial v}{\partial v^*} = \frac{\partial \hat{u}^*}{\partial v^*},$$

$$\frac{\partial \hat{v}}{\partial u}\,\frac{\partial u}{\partial u^*} + \frac{\partial \hat{v}}{\partial v}\,\frac{\partial v}{\partial u^*} = \frac{\partial \hat{v}^*}{\partial u^*}, \qquad \frac{\partial \hat{v}}{\partial u}\,\frac{\partial u}{\partial v^*} + \frac{\partial \hat{v}}{\partial v}\,\frac{\partial v}{\partial v^*} = \frac{\partial \hat{v}^*}{\partial v^*}$$

hold, we immediately see that $d\hat{u} = d\hat{u}^*$ and $d\hat{v} = d\hat{v}^*$, i.e. that $d\hat{\mathbf{r}} = d\hat{\mathbf{r}}^*$. This shows that $d\hat{\mathbf{r}}$ is independent of the choice of parametrization γ. Similarly it can be shown (do it!) that the vector is also independent of the choice of parametrization $\hat{\gamma}$. Consequently, *mapping* (23) *is defined correctly*.

Definition 4. Mapping (23) is called the *differential* of a mapping f at a point p (or the *principal linear part* of f) and is denoted by $(df)_p$ (or $\mathbf{T}_pf$).

Problem 4. An arbitrary curve (5) on $\mathscr{X}$ is carried by f into a curve

$$(25) \qquad \hat{u} = \hat{u}\,(u\,(t),\ v\,(t)), \quad \hat{v} = \hat{v}\,(u\,(t),\ v\,(t)), \quad t \in I$$

on $\hat{\mathcal{X}}$. Let for $t = t_0$ the curve (5) pass through p_0 of $\mathcal{X}$. Show that the differential $(df)_{p_0}$ of f carries the tangent vector to the curve (5) at p_0 into the tangent vector to the curve (25) at $f(p_0)$.

A mapping $f\colon \mathcal{X} \to \hat{\mathcal{X}}$ is called a *diffeomorphism* if so is the mapping $\bar{f}$. (This definition is correct.) Functions (22) which specify a diffeomorphism have, as is known from calculus, the property that their Jacobian

$$\begin{vmatrix} \dfrac{\partial \hat{u}}{\partial u} & \dfrac{\partial \hat{u}}{\partial v} \\ \dfrac{\partial \hat{v}}{\partial u} & \dfrac{\partial \hat{v}}{\partial v} \end{vmatrix}$$

is nonzero. This means (see formulas (24)) that the *differential* $(df)_p$ *of the diffeomorphism* f *at every point* $p \in \mathcal{X}$ *is an isomorphism of a vector space* $\mathbf{T}_p\mathcal{X}$ *onto* $\mathbf{T}_{\hat{p}}\hat{\mathcal{X}}$.

For $\hat{U} = U$, the formulas $\hat{u} = u$, $\hat{v} = v$ obviously define a diffeomorphism. This diffeomorphism is a mapping which *preserves the coordinates of vectors in another system of coordinates.*

It is interesting to note that *for any diffeomorphism* $f\colon \mathcal{X} \to \hat{\mathcal{X}}$ *and any parametrization* $\gamma\colon U \to \mathcal{A}$ *of* $\mathcal{X}$ *there is a parametrization* $\gamma^*\colon U \to \mathcal{A}$ *of* $\hat{\mathcal{X}}$ *such that in* γ *and* γ^* *the diffeomorphism* f *is a mapping which preserves the coordinates of vectors in another system of coordinates.* Indeed, let $\hat{\gamma}\colon \hat{U} \to \mathcal{A}$ be an arbitrary parametrization of $\hat{\mathcal{X}}$ and let the mapping f be represented by the diffeomorphism $\bar{f}\colon U \to \hat{U}$ in γ and $\hat{\gamma}$. Consider a composite mapping $\gamma^* = \hat{\gamma} \circ \bar{f}$. Since $\bar{f}$ is a diffeomorphism, this mapping is a parametrization of $\hat{\mathcal{X}}$ equivalent to $\hat{\gamma}$. Since $f \circ \gamma = \hat{\gamma} \circ \bar{f} = \gamma^* \circ \mathrm{id}$, where id is the identity mapping, in γ and γ^* the deffeomorphism f is represented by id and hence preserves the coordinates of vectors in another system of coordinates. □

As we shall see later, this property of diffeomorphisms often significantly facilitates their study.

Now we shall assume a space $\mathscr{A}$ containing a given surface $\mathscr{X}$ to be *Euclidean*. Then a Euclidean structure arises in every tangent subspace $\mathbf{T}_p\mathscr{X}$, the square of the length of an arbitrary vector (18) of that subspace being expressed as follows:

$$(26) \qquad d\mathbf{r}^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2,$$

where

$$E = \mathbf{r}_u^2, \quad F = \mathbf{r}_u\mathbf{r}_v, \quad G = \mathbf{r}_v^2$$

are metric coefficients of the basis $\mathbf{r}_u$, $\mathbf{r}_v$.

Definition 5. A quadratic form (26) of variables du and dv is called the *first quadratic form* of a surface $\mathscr{X}$. It is usually denoted by $\mathbf{I}$ or $\mathbf{I}\,(d\mathbf{r})$. Thus by definition

$$\mathbf{I} = d\mathbf{r}^2.$$

Note that the coefficients E, F, and G of the first quadratic form depend on $p \in \mathscr{X}$ and in the coordinates u and v are the smooth functions

$$E = E\,(u, v), \quad F = F\,(u, v), \quad G = G\,(u, v)$$

of u and v.

Remark 2. The expression $E\,du^2 + 2F\,du\,dv + G\,dv^2$ is usually understood as a quadratic form. From the modern point of view it should be treated as a quadratic functional on a space $\mathbf{T}_p\mathscr{X}$ whose value on vector (18) is $E\;du^2 + 2F\,du\,dv + G\,dv^2$.

The length $|\mathbf{r}'\,(t)|$ of the tangent vector $\mathbf{r}'\,(t)$ of curve (6) on $\mathscr{X}$ can be expressed in view of (16) as follows:

$$|\mathbf{r}'\,(t)| = \sqrt{\mathbf{r}'\,(t)^2} = \sqrt{(u'\,(t)\,\mathbf{r}_u + v'\,(t)\,\mathbf{r}_v)^2}$$
$$= \sqrt{Eu'\,(t)^2 + 2F\,u'(t)\,v'\,(t) + G\,v'(t)^2},$$

where, of course, E, F, and G are regarded as functions of t:

$$E = E\,(u\,(t),\ v\,(t)), \quad F = F\,(u\,(t),\ v\,(t)),$$
$$G = G\,(u\,(t),\ v\,(t)).$$

For the length s of curve (6) this yields the formula

$$(27) \qquad s = \int_a^b \sqrt{E\,u'(t)^2 + 2F\,u'(t)\,v'(t) + G\,v'(t)^2}\,dt$$

which can be rewritten in the following handy mnemonic form:

$$s = \int_L \sqrt{E\,du^2 + 2F\,du\,dv + G\,dv^2},$$

where L designates curve (6).

The formula

$$(28) \qquad ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2$$

is of a still more conventional form (in shorter notation $ds^2 = \mathrm{I}\,(d\mathbf{r})$ or $ds^2 = d\mathbf{r}^2$) which means that the *first quadratic form* **I** *specifies the square* ds^2 *of the element of length.*

[It should be remembered that both the formulation and formula (28) are conventional in character, they serve only as shortened expressions of formula (27).]

By the general rules of linear algebra, for the angle θ between two tangent vectors

$$d\mathbf{r} = \mathbf{r}_u du + \mathbf{r}_v\,dv \quad \text{and} \quad \delta\mathbf{r} = \mathbf{r}_u \delta u + \mathbf{r}_v \delta v$$

we have

$$\cos\theta = \frac{d\mathbf{r}\,\delta\mathbf{r}}{|\,d\mathbf{r}\,|\;|\,\delta\mathbf{r}\,|},$$

i.e.

$$(29) \quad \cos\theta = \frac{E\,du\,\delta u + F\,(du\,\delta v + \delta u\,dv) + G\,dv\,\delta v}{\sqrt{E\,du^2 + 2F\,du\,dv + G\,dv^2}\,\sqrt{E\,\delta u^2 + 2F\,\delta u\,\delta v + G\,\delta v^2}}$$

which can be arbitrarily written in the following mnemonic form:

$$\cos\theta = \frac{\mathrm{I}\,(d,\,\delta)}{\sqrt{\mathrm{I}\,(d)}\,\sqrt{\mathrm{I}\,(\delta)}}.$$

The angle between two tangent vectors

$$\mathbf{r}'\,(t_0) = u'\,(t_0)\,\mathbf{r}_{u_0} + v'\,(t_0)\mathbf{r}_{v_0} \quad \text{and}$$

$$\mathbf{r}_1'\,(t_0) = u_1'\,(t_0)\mathbf{r}_{1u_0} + v_1'\,(t_0)\,\mathbf{r}_{1v_0}$$

at p_0 is called the *angle between two curves*

$$\mathbf{r} = \mathbf{r}(u(t), \quad v(t)) \quad \text{and} \quad \mathbf{r}_1 = \mathbf{r}_1(u_1(t), \quad v_1(t))$$

on $\mathscr{X}$, which for $t = t_0$ pass through the same point p_0. We have

$$\cos\theta = \frac{\mathbf{r}'(t_0)\,\mathbf{r}_1'(t_0)}{|\mathbf{r}'(t_0)|\cdot|\mathbf{r}_1'(t_0)|},$$

where

$$\begin{aligned}\mathbf{r}'(t_0)\,\mathbf{r}_1'(t_0)&\\ &= E_0u'(t_0)\,u_1'(t_0) + F_0(u'(t_0)\,v_1'(t_0) + u_1'(t_0)\,v'(t_0))\\ &\quad + G_0v'(t_0)\,v_1'(t_0),\\ |\mathbf{r}'(t_0)| &= \sqrt{E_0u'(t_0)^2 + 2F_0u'(t_0)\,v'(t_0) + G_0v'(t_0)^2},\\ |\mathbf{r}_1'(t_0)| &= \sqrt{E_0u_1'(t_0)^2 + 2F_0u_1'(t_0)\,v_1'(t_0) + G_0v_1'(t_0)^2}\end{aligned}$$

and

$$\begin{aligned}E_0 &= E(u_0, v_0) = E(u(t_0), v(t_0)),\\ F_0 &= F(u_0, v_0) = F(u(t_0), v(t_0)),\\ G_0 &= G(u_0, v_0) = G(u(t_0), v(t_0))\end{aligned}$$

are the values of the coefficients of the first quadratic form at p_0.

In particular, for the cosine of the angle between two coordinate lines u = const and v = const, we get

$$\cos\theta = \frac{F}{\sqrt{E}\sqrt{G}},$$

from which it follows that the *coordinate lines* u = const *and* v = const *are orthogonal if and only if* $F = 0$.

Thus *we can compute the lengths of curves on a surface and the angles between them knowing only the first quadratic form of that surface*, i.e. the Euclidean structures on all tangent planes.

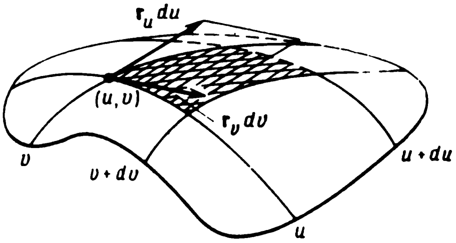
Known in calculus is the concept of *area* of an arbitrary part D of a surface $\mathscr{X}$ as the limit of areas that approximate that part of polyhedral surfaces. This limit (if any) is expressed by the integral

$$\iint\limits_D \sqrt{\Gamma(\mathbf{r}_u, \mathbf{r}_v)}\,du\,dv,$$

where

$$\Gamma(\mathbf{r}_u, \mathbf{r}_v) = \begin{vmatrix} \mathbf{r}_u^2 & \mathbf{r}_u\mathbf{r}_v \\ \mathbf{r}_u\mathbf{r}_v & \mathbf{r}_v^2 \end{vmatrix} = \begin{vmatrix} E & F \\ F & G \end{vmatrix} = EG - F^2$$

is the *Gram determinant* or *Gramian* of vectors $\mathbf{r}_u$ and $\mathbf{r}_v$. An intuitive idea of this formula is that the area of an infinitesimal curvilinear parallelogram of a coordinate network with vertex at a point (u, v) and sides du and dv is approximately equal to that of a parallelogram in a tangent plane with sides $\mathbf{r}_u du$ and $\mathbf{r}_v dv$ (see the figure). On the other hand, it is easy to prove (see Lemma 1 of Lecture II.26) that this area is equal to $\sqrt{\Gamma(\mathbf{r}_u, \mathbf{r}_v)}\, du\, dv$. In calculus it is said that an *element of a surface area is equal to* $\sqrt{EG - F^2}\, du\, dv$.



Thus the first quadratic form allows us to compute areas on $\mathscr{X}$.

For brevity a linear isomorphism of Euclidean vector spaces is called an isometry. By definition (see Definition 7 of Lecture II.14) the linear isomorphism $\varphi\colon \mathscr{V} \to \mathscr{V}_1$ is an isometry if

$$\mathbf{xy} = \varphi\mathbf{x} \cdot \varphi\mathbf{y} \tag{30}$$

for any vectors $\mathbf{x}, \mathbf{y} \in \mathscr{V}$. For condition (30) to hold it suffices that $\mathbf{x} = \mathbf{y}$, i.e. a *linear mapping* $\varphi\colon \mathscr{V} \to \mathscr{V}_1$ *having the property that*

$$\mathbf{x}^2 = (\varphi\mathbf{x})^2 \text{ for any vector } \mathbf{x} \in \mathscr{V} \tag{31}$$

is an isometry. (To prove this it suffices to apply relation (31) to the vector $\mathbf{x} + \mathbf{y}$.)

In coordinates the mapping φ is written by the following linear formulas

$$y^i = a^i_j x^j$$

and the scalar squares $\mathbf{x}^2$, $\mathbf{x} \in \mathscr{V}$, and $\mathbf{y}^2$, $\mathbf{y} \in \mathscr{V}_1$, by the quadratic forms

$$\mathbf{x}^2 = g_{ij}x^i x^j, \quad \mathbf{y}^2 = h_{ij}y^i y^j.$$

For this notation equation (31) implies that substituting expressions $y^i = a^i_j x^j$ in the form $h_{ij}y^i y^j$ yields $g_{ij}x^i x^j$, i.e. that there are

$$g_{ij} = h_{kl}a^k_i{}^a a^l_j.$$

In particular, for Euclidean two-dimensional spaces $\mathbf{T}_p\mathscr{X}$, $\mathbf{T}_{\hat{p}}\hat{\mathscr{X}}$ and the differential $(df)_p\colon \mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\hat{p}}\hat{\mathscr{X}}$, $\hat{p} = f(p)$, of the diffeomorphism $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ the fact that conditions (31) hold at all points $p \in \mathscr{X}$ implies that substituting in the first quadratic form $\hat{E}\,d\hat{u}^2 + 2\hat{F}\,d\hat{u}\,d\hat{v} + \hat{G}\,d\hat{v}^2$ of $\hat{\mathscr{X}}$ expressions (24) for $d\hat{u}$ and $d\hat{v}$ (and in its coefficients $\hat{E}$, $\hat{F}$, $\hat{G}$, expressions (22) for $\hat{u}$ and $\hat{v}$) yields the first quadratic form $E\,du^2 + 2F\,du\,dv + G\,dv^2$ of $\mathscr{X}$, which is identical in u and v to

$$\begin{aligned}
&E(u, v)\\
&= \hat{E}(\hat{u}, \hat{v})\left(\frac{\partial\hat{u}}{\partial u}\right)^2 + 2\hat{F}(\hat{u}, \hat{v})\frac{\partial\hat{u}}{\partial u}\,\frac{\partial\hat{v}}{\partial u} + \hat{G}(\hat{u}, \hat{v})\left(\frac{\partial\hat{v}}{\partial u}\right)^2,\\
&F(u, v)\\
&= \hat{E}(\hat{u}, \hat{v})\frac{\partial\hat{u}}{\partial u}\,\frac{\partial\hat{u}}{\partial v} + \hat{F}(\hat{u}, \hat{v})\left(\frac{\partial\hat{u}}{\partial u}\,\frac{\partial\hat{v}}{\partial v} + \frac{\partial\hat{u}}{\partial v}\,\frac{\partial\hat{v}}{\partial u}\right)\\
&+ \hat{G}(\hat{u}, \hat{v})\frac{\partial\hat{v}}{\partial u}\,\frac{\partial\hat{v}}{\partial v},\\
&(32)\\
&G(u, v)\\
&= \hat{E}(\hat{u}, \hat{v})\left(\frac{\partial\hat{u}}{\partial v}\right)^2 + 2\hat{F}(\hat{u}, \hat{v})\frac{\partial\hat{u}}{\partial v}\,\frac{\partial\hat{v}}{\partial v} + \hat{G}(\hat{u}, \hat{v})\left(\frac{\partial\hat{v}}{\partial v}\right)^2.
\end{aligned}$$

Thus *for a diffeomorphism* $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ *linear mappings*

$$(df)_p\colon \mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\hat{p}}\hat{\mathscr{X}}, \quad \hat{p} = f(p)$$

are isometries for all points $p \in \mathscr{X}$ *if and only if equations* (32) *hold identically in* u *and* v.

For the diffeomorphism $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ preserving the coordinates of vectors in another system of coordinates this condition implies that in the coordinates under consideration the *first quadratic forms of* $\mathscr{X}$ *and* $\hat{\mathscr{X}}$ *coincide* (or to be exact, they differ only in the notation of variables).

Definition 6. A diffeomorphism $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ is said to be an *isometric mapping* of a surface $\mathscr{X}$ onto $\hat{\mathscr{X}}$ (or simply an *isometry*) if for any curve

$$u = u(t), \quad v = v(t), \quad a \leqslant t \leqslant b, \tag{33}$$

on $\mathscr{X}$ its image

$$\hat{u} = \hat{u}(t), \quad \hat{v} = \hat{v}(t), \quad a \leqslant t \leqslant b$$

for $\hat{\mathscr{X}}$, where $\hat{u}(t) = \hat{u}(u(t), v(t))$ and $\hat{v}(t) = \hat{v}(u(t), v(t))$, has the same length, i.e. if

$$\int_a^b \sqrt{E(t)u'(t)^2 + 2F(t)u'(t)v'(t) + G(t)v'(t)^2}\,dt$$

$$= \int_a^b \sqrt{\hat{E}(t)\hat{u}'(t)^2 + 2\hat{F}(t)\hat{u}'(t)\hat{v}'(t) + \hat{G}(t)\hat{v}'(t)^2}\,dt,$$

where

$$E(t) = E(u(t), v(t)), \quad F(t) = F(u(t), v(t)),$$
$$G(t) = G(u(t), v(t))$$

and similarly

$$\hat{E}(t) = \hat{E}(\hat{u}(t), \hat{v}(t)), \quad \hat{F}(t) = \hat{F}(\hat{u}(t), \hat{v}(t)),$$
$$\hat{G}(t) = \hat{G}(\hat{u}(t), \hat{v}(t)).$$

Surfaces for which there is at least one isometric mapping $\mathscr{X} \to \hat{\mathscr{X}}$ are called *isometric.*

Proposition 1. *A diffeomorphism $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ is an isometry if and only if for any point $p \in \mathscr{X}$ so is its differential*

$$(df)_p\colon \mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\hat{p}}\hat{\mathscr{X}}, \quad \hat{p} = f(p).$$

Proof. According to the above remarks we may assume without loss of generality that the diffeomorphism f acts as a function defined by equating coordinates. Then for any curve (33) on $\mathscr{X}$ its image on $\hat{\mathscr{X}}$ under the diffeomorphism f has *the same* parametric equations (33) and the statement that the differential $(df)_p$ of f is an isometry at every point $p \in \mathscr{X}$ will mean that the first quadratic forms of $\mathscr{X}$ and $\hat{\mathscr{X}}$ differ only in the notation of variables. Formulas (27) will therefore be identical for both curves and hence the lengths of these curves will be identical. Consequently, the diffeomorphism f will be an isometry.

Conversely, let a diffeomorphism $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ preserving the coordinates of vectors in another system of coordinates be an isometry. This means that for any smooth functions $u = u(t)$, $v = v(t)$, $a \leqslant t \leqslant b$ with the property that $(u(t), v(t))$, $\in U$, $a \leqslant t \leqslant b$, we have

$$\int_a^b \sqrt{E(t)\,u'(t)^2 + 2F(t)\,u'(t)\,v'(t) + G(t)\,v'(t)^2}\,dt$$

$$= \int_a^b \sqrt{\hat{E}(t)\,u'(t)^2 + 2\hat{F}(t)\,u'(t)\,v'(t) + \hat{G}(t)\,v'(t^2)}\,dt.$$

Differentiating this identity with respect to b (and replacing b by t) we obtain, after squaring, the identity

$$E(t)\,u'(t)^2 + 2F(t)\,u'(t)\,v'(t) + G(t)\,v'(t)^2$$

$$= \hat{E}(t)\,u'(t)^2 + 2\hat{E}(t)\,u'(t)\,v'(t) + \hat{G}(t)\,v'(t)^2.$$

In particular, this identity must hold for linear functions of the form

$$u(t) = u_0 + \alpha t, \quad v(t) = v_0 + \beta t, \quad |t| < \varepsilon,$$

where (u_0, v_0) is an arbitrary point of a domain U, α and β are arbitrary numbers, and $\varepsilon > 0$ is a sufficiently small positive number. But in this case, after substituting $t = 0$ it becomes

$$E_0\alpha^2 + 2F_0\alpha\beta + G_0\beta^2 = \hat{E}_0\alpha^2 + 2\hat{F}_0\alpha\beta + \hat{G}_0\beta^2,$$

where $E_0, \ldots, \hat{G}_0$ are the values of the functions $E, \ldots, \hat{G}$ at (u_0, v_0), and is therefore possible, in view of α and β being arbitrary numbers, if and only if $E_0 = \hat{E}_0$, $F_0 = \hat{F}_0$ and $G_0 = \hat{G}_0$, i.e. if $E = \hat{E}$, $F = \hat{F}$ and $G = \hat{G}$ everywhere in U. Hence linear mappings $(df)_p$ are isometries. ☐

Corollary 1. *On isometric surfaces the corresponding curves intersect at the same angles, and the corresponding domains have the same areas.* ☐

Corollary 2. *Two surfaces are isometric if and only if local coordinates can be chosen on them in which the first quadratic forms of those surfaces coincide.* ☐

Of course, this criterion for isometry is extremely ineffective (how can we guess if there are local coordinates it provides for?). Our final goal (to be achieved in Lecture 5) is to make this criterion effective, but to do this we shall have to go quite a long way.

Assuming a surface to be made from a flexible but unstretchable material and bending it arbitrarily we shall not change the lengths of curves on the surface and hence get an isometric surface. Based on this graphical representation the founders of the theory of surfaces in the 19th century called isometries the *bendings*. This terminology has partly survived till the present day, but nowadays bendings are usually understood in a narrower sense, as isometries that can be connected with the identity transformation by a continuous family of isometries. All mathematicians were sure for a long time that in a local situation, i.e. in a sufficiently small neighbourhood of an arbitrary point, any isometry is a bending in this sense. Relatively recently, however, N. V. Efimov has shown that this is wrong by constructing an appropriate counterexample.

Suppose a surface is inhabited by intelligent creatures that can measure lengths, areas, and angles, but cannot reach out of the space. Then with any bending of the surface all their geometry will remain the same and they will simply fail to notice this bending. For this reason isometric surfaces are said to have the same *intrinsic geometry*.

Here is one important example of an intrinsic and geometric construction.

According to the general results of linear algebra (see Lecture II.14) for any Euclidean space $\mathscr{V}$ with metric tensor g_{ij} the conjugate space $\mathscr{V}^*$ is a Euclidean space with metric tensor g^{ij} whose components form a matrix $\| g^{ij} \|$ inverse to the matrix $\| g_{ij} \|$ of the components of g_{ij}. For any covector $\xi = (\xi_1, \ldots, \xi_n)$ in $\mathscr{V}^*$ its length $| \xi |$ is thus defined whose square is expressed as follows:

$$| \xi | = g^{ij}\xi_i\xi_j.$$

In our case, for a two-dimensional Euclidean space $\mathbf{T}_p\mathscr{X}$ the matrix $\| g_{ij} \|$ is of the form

$$\begin{Vmatrix} E & F \\ F & G \end{Vmatrix}$$

and hence $\| g^{ij} \|$ is of the form

$$\frac{1}{EG-F^2}\begin{Vmatrix} G & -F \\ -F & E \end{Vmatrix}.$$

For the length $| \xi |$ of an arbitrary covector ξ on $\mathbf{T}_p\mathscr{X}$ we have

$$|\xi|^2 = \frac{G\xi^2 - 2F\xi\eta + E\eta^2}{EG-F^2},$$

where ξ, η are the coördinates of that covector.

An example of the covector ξ is the gradient $\left(\frac{\partial\varphi}{\partial u}, \frac{\partial\varphi}{\partial v}\right)$ of an arbitrary smooth function $\varphi = \varphi(u, v)$ on $\mathscr{X}$. The square of the length of that covector is called *Beltrami's differential operator* of the first kind of the function φ and is denoted by $\Delta_1\varphi$. Thus by definition

$$\Delta_1\varphi = \frac{G\varphi_u^2 - 2F\varphi_u\varphi_v + E\varphi_v^2}{EG-F^2}.$$

This construction is invariant, i.e. independent of the choice of parametrization, and for any isometry $f\colon \mathscr{X} \to \hat{\mathscr{X}}$ we have

$$\Delta_1 (\varphi \circ f) = \Delta_1 \varphi \circ f \tag{34}$$

(check it!). In this sense it belongs to the intrinsic geometry of the surface $\mathscr{X}$.

Consider in conclusion some examples of computing the first quadratic form of surfaces in a three-dimensional Euclidean space. In these examples the surfaces are not elementary as a rule. But they are easily reduced to elementary ones by cuttings and restrictions of the domains of definition of parametrizations.

Example 4. The plane *Oxy* has the parametric equation $\mathbf{r} = u\mathbf{i} + v\mathbf{j}$ in coordinates $u = x$ and $v = y$. Therefore $\mathbf{r}_u = \mathbf{i}$, $\mathbf{r}_v = \mathbf{j}$ and hence $E = 1$, $F = 0$, $G = 1$, i.e. for a plane,

$$\mathrm{I} = du^2 + dv^2. \tag{35}$$

(The result is easy to foresee without any computations.)

An open subset of a plane (considered as a surface in space) also has the same quadratic form.

Example 5. For a circular cylinder

$$\mathbf{r} = R \cos u \cdot \mathbf{i} + R \sin u \cdot \mathbf{j} + v \cdot \mathbf{k}$$

we have $\mathbf{r}_u = -R \sin u \cdot \mathbf{i} + R \cos u \cdot \mathbf{j}$ and $\mathbf{r}_v = \mathbf{k}$. Therefore

$$E = \mathbf{r}_u^2 = R^2, \quad F = \mathbf{r}_u \mathbf{r}_v = 0, \quad G = \mathbf{r}_v^2 = 1,$$

i.e. for a cylinder

$$\mathrm{I} = R^2\, du^2 + dv^2.$$

Introducing a new coordinate $u_1 = Ru$ (and denoting u_1 by u) we transform this to form (35).

Thus, there are coordinates in which the first quadratic form of a plane and that of a cylinder coincide! This does not mean yet that the plane and the cylinder (of course, cut in; see Example 1 above) are isometric, since for a cylinder the coordinates range over only some region in $\mathbb{R}^2$ and so the cut-in cylinder is only isometric to a part of the plane. We express this by saying that the cylinder and the plane are *locally isometric*.

An isometric mapping of a cut-in cylinder onto a flat region is produced graphically by gradually unbending it.

Example 6. For a surface of revolution

$$\mathbf{r} = x(v)\cos u\cdot\mathbf{i} + x(v)\sin u\cdot\mathbf{j} + z(v)\cdot\mathbf{k}$$

we have

$$\mathbf{r}_u = -x(v)\sin u\cdot\mathbf{i} + x(v)\cos u\cdot\mathbf{j},$$

$$\mathbf{r}_v = x'(v)\cos u\cdot\mathbf{i} + x'(v)\sin u\cdot\mathbf{j} + z'(v)\,\mathbf{k}.$$

Hence

$$E = x(v)^2\sin^2 u + x(v)^2\cos^2 u = x(v)^2,$$

$$F = -x(v)\sin u\cdot x'(v)\cos u + x(v)\cos u\cdot x'(v)\sin u = 0,$$

$$G = x'(v)^2\cos^2 u + x'(v)^2\sin^2 u + z'(v)^2 = x'(v)^2 + z'(v)^2,$$

so that for the surface of revolution

$$\mathrm{I} = x(v)^2 du^2 + (x'(v)^2 + z'(v)^2)\,dv^2.$$

It is graphically obvious that the meridians and parallels of any surface of revolution are orthogonal. The equation $F = 0$ could therefore be foreseen without any computations as well.

When the profile $x = x(v)$, $z = z(v)$ of a surface of revolution is referred to the natural parameter $v = s$ (and therefore $x'(v)^2 + z'(v)^2 = 1$) the form I is particularly simple:

$$\mathrm{I} = x(v)^2 du^2 + dv^2.$$

In particular, we see that the *first quadratic form of a sphere of radius* 1 *is of the form*

$$\mathrm{I} = \cos^2 v\,du^2 + dv^2. \tag{36}$$

Cartographic experience shows that no portion of a sphere however small can be bent into a plane. This means that no transformation of coordinates can convert form (36) into form (35). But how is this to be proved? The answer will be given in Lecture 5.

Example 7. The curve formed by a chain of uniform density hanging freely from two fixed points not in the same vertical line is called a *catenary* and a surface of revolution whose profile is a catenary is called a *catenoid.*

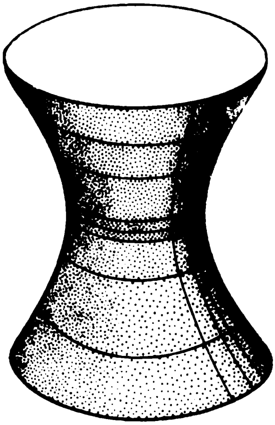
In mechanics (statics) a catenary is the graph of a hyperbolic cosine. Thus for a catenoid $x(v) = \cosh v$, $z(v) = v$ and hence

$$x(v)^2 = \cosh^2 v \text{ and } x'(v)^2 + z'(v)^2 = \sinh^2 v + 1 = \cosh^2 v.$$

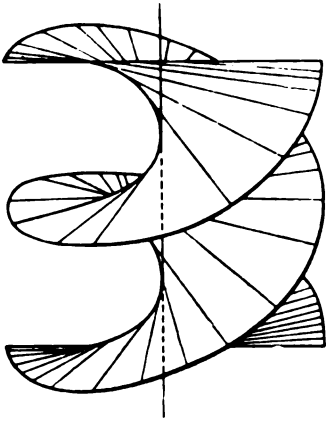
Thus for the catenoid

$$\mathrm{I} = \cosh^2 v\,(du^2 + dv^2). \tag{37}$$

Example 8. Let a straight line perpendicular to the axis Oz rotate uniformly near it remaining perpendicular to it in rotation and simultaneously ascending in helical motion (to a height proportional to the angle of rotation). The ruled surface generated by a moving straight line is called a *helicoid.* It has the form of a helical ramp for cars to drive up.



A catenoid



A helicoid

If v is the parameter on the straight line and u is the angle of rotation, then the helicoid will have the equation

$$\mathbf{r} = v \cos u \cdot \mathbf{i} + v \sin u \cdot \mathbf{j} + u \cdot \mathbf{k}.$$

Therefore

$$\mathbf{r}_u = -v \sin u \cdot \mathbf{i} + v \cos u \cdot \mathbf{j} + \mathbf{k},$$

$$\mathbf{r}_v = \cos u \cdot \mathbf{i} + \sin u \cdot \mathbf{j}$$

and hence

$$E = 1 + v^2, \quad F = 0, \quad G = 1.$$

Thus for a helicoid

$$\mathrm{I} = (1 + v^2)\, du^2 + dv^2.$$

Let us transform this form by introducing new coordinates u_1, v_1 related to the coordinates u, v as follows:

$$u = u_1, \quad v = \sinh v_1.$$

Then

$$1 + v^2 = 1 + \sinh^2 v_1 = \cosh^2 v_1,$$

$$du = du_1, \quad dv = \cosh v_1 dv_1$$

and therefore (we drop the indices in the new coordinates)

$$\mathrm{I} = \cosh^2 v\, (du^2 + dv^2),$$

which coincides with form (37).

This proves that the *catenoid and the helicoid are locally isometric* (more exactly, the catenoid cut-in along a meridian is isometric to a part $0 < u < 2\pi$ of the helicoid), there being an isometry which sends meridians of the catenoid into rectilinear generators of the helicoid.

Example 9. For an arbitrary ruled surface

$$\mathbf{r} = \boldsymbol{\rho}(u) + v\mathbf{a}(u), \tag{38}$$

where (see Example 3) $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ is a regular curve referred to the natural parameter and $\mathbf{a}(u)$ is a vector function such that $|\mathbf{a}(u)| = 1$ for all u, denoting differentiation with respect to u by a dot, we have

$$\mathbf{r}_u = \dot{\boldsymbol{\rho}} + v\dot{\mathbf{a}}, \quad \mathbf{r}_v = \mathbf{a}.$$

Since $\dot{\boldsymbol{\rho}}^2 = 1$, and $\mathbf{a}^2 = 1$ and $\mathbf{a}\dot{\mathbf{a}} = 0$, we have

$$E = 1 + 2v\dot{\boldsymbol{\rho}}\dot{\mathbf{a}} + v^2\dot{\mathbf{a}}^2, \quad F = \dot{\boldsymbol{\rho}}\mathbf{a}, \quad G = 1.$$

If, in particular, $\mathbf{a} = \dot{\boldsymbol{\rho}}$ (a surface of tangents), then $\dot{\boldsymbol{\rho}}\mathbf{a} = \mathbf{a}^2 = 1$ (i.e. $F = 1$) and $\dot{\boldsymbol{\rho}}\dot{\mathbf{a}} = 0$ and $\dot{\mathbf{a}}^2 = k^2$, where k is the curvature of the curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ (i.e. $E = 1 + k^2v^2$). Thus for a surface of tangents

$$\mathrm{I} = (1 + k^2v^2)\,du^2 + 2du\,dv + dv^2. \tag{39}$$

But if $\mathbf{a}(u)$ is the binormal vector of the curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$, then $\dot{\boldsymbol{\rho}}\mathbf{a} = 0$, $\dot{\boldsymbol{\rho}}\dot{\mathbf{a}} = 0$ and $\dot{\mathbf{a}}^2 = \varkappa^2$, where $\varkappa$ is the torsion of the curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$. Hence for a surface of binormals

$$\mathrm{I} = (1 + \varkappa^2v^2)\,du^2 + dv^2.$$

We thus see that the first quadratic form of a surface of tangents depends only on the curvature of a given curve and that the first quadratic form of a surface of binormals depends only on the torsion of the surface.

For surfaces of tangents this implies that *every surface of tangents is locally isometric to a plane.* Indeed, consider a plane curve with the same curvature $k = k(u)$ (such a curve exists by virtue of Theorem 1 of Lecture 2). The first quadratic form of the surface of tangents for that curve is of form (39). But, on the other hand, it is clear that a surface of tangents of a plane curve is, outside its singular points, a domain in the plane. There exists therefore a change of coordinates transforming the first quadratic form $dx^2 + dy^2$ of the plane into form (39). (This change of coordinates has the form

$$x = x(u) + x'(u)\,v, \quad y = y(u) + y'(u)\,v,$$

where $x(u)$ and $y(u)$ are functions such that $x'(u)^2 + y'(u)^2 = 1$ and $x''(u)^2 + y''(u)^2 = k(u)^2$.) □

This isometry can be carried out by continuous bending, gradually deforming the curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ into a plane curve.

For this reason surfaces of tangents are called *developable surfaces* (or developables) (development into a plane is meant).

If $\mathbf{a}(u) = \boldsymbol{\rho}(u)$, then surface (38) is a cone with vertex at the origin (and the curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ is the intersection of the cone and the unit sphere $|\boldsymbol{\rho}| = 1$). In this case we have

$$\dot{\boldsymbol{\rho}}\dot{\mathbf{a}} = \boldsymbol{\rho}^2 = 1, \quad \dot{\mathbf{a}}^2 = 1, \quad \dot{\boldsymbol{\rho}}\mathbf{a} = 0,$$

from which for the first quadratic form we get

$$\mathrm{I} = (1 + v)^2 du^2 + dv^2.$$

Here the change of coordinates $(u, v) \mapsto (u, 1 + v)$ suggests itself, which converts the last form into a slightly simpler form

$$\mathrm{I} = v^2 du^2 + dv^2. \tag{40}$$

Now let us introduce new coordinates

$$x = v \cos u, \quad y = v \sin u.$$

Then

$$dx = -v \sin u \, du + \cos u \, dv,$$

$$dy = v \cos u \, du + \sin u \, dv$$

and hence

$$dx^2 + dy^2 = v^2 du^2 + dv^2.$$

This proves that the *cone is also locally isometric to a plane* (more exactly, each nappe of a cone cut-through along a generator is isometric to some plane sector; the fact is intuitively obvious). For this reason cones are also reckoned among developables.

Notice that form (10) is nothing but the first quadratic form of a plane referred to polar coordinates $r = v$ and $\varphi = u$.

Finally, if the vector $\mathbf{a}(u)$ is constant (and therefore $\dot{\mathbf{a}} = 0$), then surface (38) is a cylinder. We may assume without loss of generality that the directrix $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ of the cylinder is a plane curve whose plane is orthogonal to the vector $\mathbf{a}$ (and hence $\boldsymbol{\rho}\mathbf{a} = 0$ and $\dot{\boldsymbol{\rho}}\mathbf{a} = 0$). Therefore, as for a circular cylinder (Example 3),

$$\mathrm{I} = du^2 + dv^2.$$

For this reason all cylinders are also reckoned among developables.

In Lecture 4 we shall show that among ruled surfaces only developables (i.e. cylinders, cones, and surfaces of tangents) are locally isometric to a plane, and in Lecture 5 it will be shown that developables exhaust all the surfaces of a three-dimensional space which are locally isometric to a plane.

Lecture 4

Normal vector · Surface as the graph of a function · Normal sections · The second quadratic form of a surface · The Dupin indicatrix · Principal, total and mean curvatures · The second quadratic form of a graph · Ruled surfaces of zero curvature · Surfaces of revolution

---

In this lecture we shall take a closer look at surfaces in a three-dimensional oriented Euclidean space. As a rule we shall study surfaces only locally, i.e. in sufficiently small neighbourhoods of their points. We may therefore assume, without loss of generality, all surfaces to be elementary.

So let $\mathscr{X}$ be an arbitrary elementary surface in a three-dimensional oriented Euclidean space $\mathscr{A}$ and let $\mathbf{r} = \mathbf{r}(u, v)$ be its arbitrary parametrization. Then at every point $p \in \mathscr{X}$ there is a single vector $\mathbf{n}$ of unit length which is perpendicular to the tangent plane and forms together with vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ a positively oriented basis for the space. This vector is given as follows:

$$\mathbf{n} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|}$$

and is called the *normal vector* to $\mathscr{X}$ at p (and the basis $\mathbf{r}_u, \mathbf{r}_v, \mathbf{n}$ is called the *moving basis* of the surface). See the figure on page 68.

By definition the square of the length $|\mathbf{r}_u \times \mathbf{r}_v|^2$ of a vector product is equal to the area of the parallelogram constructed on $\mathbf{r}_u$ and $\mathbf{r}_v$, i.e. (see Lecture 3) to the

Gramian $\Gamma(\mathbf{r}_u, \mathbf{r}_v) = EG - F^2$ of $\mathbf{r}_u$ and $\mathbf{r}_v$. Hence

$$\mathbf{n} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{\sqrt{EG - F^2}} .$$

It is this formula that is usually used to compute **n**.

Choosing in $\mathscr{A}$ rectangular coordinates x, y, and z with the origin O at a point p and the Oz-axis directed along the vector **n**, consider a Jacobian matrix

$$\begin{Vmatrix} x_u & y_u & z_u \\ x_v & y_v & z_v \end{Vmatrix}$$

whose rows are the coordinates of $\mathbf{r}_u$ and $\mathbf{r}_v$. The condition imposed on Oz implies, in particular, that

$$\begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix} = 1.$$

Therefore according to the **inverse mapping theorem** known from calculus, in some neighbourhood of the point (0, 0) the coordinates x and y can be expressed in terms of the coordinates u and v. Substituting these expressions in the vector function $\mathbf{r} = \mathbf{r}(u, v)$ we obtain a parametrization of the surface $\mathscr{X}$ (or more precisely some part of the surface, which contains the point p) of the form

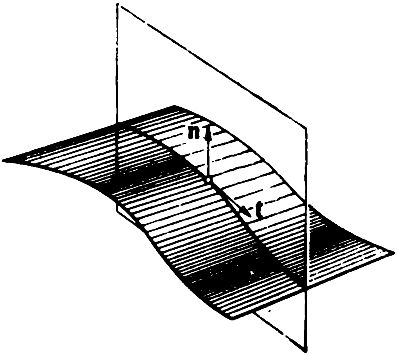
$$\begin{aligned} x &= x, \\ y &= y, \\ z &= z(x, y). \end{aligned}$$

By definition this means that *locally* (in a neighbourhood of p) *the surface* $\mathscr{X}$ *is the graph of the function* $z = z(x, y)$.

Every plane Π passing through the Oz-axis of the coordinate system involved is called the *normal plane* of $\mathscr{X}$ at p and its intersection $\Pi \cap \mathscr{X}$ with the surface $\mathscr{X}$ is the *normal section* of $\mathscr{X}$ at p. The direction bivector of every normal plane Π is of the form $\mathbf{t} \wedge \mathbf{n}$, where $\mathbf{t}$ is some nonzero tangent vector of $\mathscr{X}$, defined up to proportionality, at p (a vector of the space $\mathbf{T}_p\mathscr{X}$). The vector $\mathbf{t}$ uniquely defines the normal plane Π and we shall denote the plane by Π_t.

If x, y, z are chosen so that $\mathbf{t}$ is directed along Ox, then Π_t is the coordinate plane Oxz and the normal section $\mathscr{L}_t = \Pi \cap \mathscr{X}$ has the equation $z = z(x, 0)$ (in coordinates x, z in the plane Oxz). This proves that (in some neighbourhood of p) every normal section is locally a graph and hence a simple singular arc.

A tangent to the graph $z = z(x, 0)$ is, of course, in the tangent plane to the surface $z = z(x, y)$ and is at the same time in the normal plane. It is therefore directed along $\mathbf{t}$, i.e. the tangent vector to the normal section at p is proportional to $\mathbf{t}$.



Normal section

Since $\mathbf{t}$ is only defined by the normal section only up to proportionality, without loss of generality we may assume it to be a unit vector. Referring the normal section to the natural parameter we may therefore consider the unit vector $\mathbf{t}$ as the tangent vector to the normal section at p.

Thinking of the normal section as a curve on Π_t we may speak of the relative curvature k_{rel} of the curve at p with respect to the orientation of Π_t given by the bivector $\mathbf{t} \wedge \mathbf{n}$. Denoting this curvature by $k(\mathbf{t})$ we thus define some function $\mathbf{t} \mapsto k(\mathbf{t})$ on unit vectors of the tangent space $\mathbf{T}_p\mathscr{X}$. We find the expression for that function in terms of the coordinates of $\mathbf{t}$.

Let

$$u = u(s), \quad v = v(s), \quad |s| < s_0, \tag{1}$$

be the parametric equation of the normal section $\mathscr{L}_t$ on $\mathscr{X}$ where s is the natural parameter counted off from p. As a curve in space the section $\mathscr{L}_t$ has the vector parametric equation $\mathbf{r} = \mathbf{r}(u(s), v(s))$ and hence the tangent vector of the curve, $\dot{\mathbf{r}}$, is expressed as follows:

$$\dot{\mathbf{r}} = \mathbf{r}_u \dot{u} + \mathbf{r}_v \dot{v}, \tag{2}$$

where as ever $\dot{u}$ and $\dot{v}$ are the derivatives of functions (1) with respect to s.

The derivative $\ddot{\mathbf{r}}$ of vector (2) is orthogonal to it, and parallel to the plane Π_t for $s = 0$. For $s = 0$ therefore it is proportional to $\mathbf{n}$. The corresponding proportionality factor (equal to the scalar product $\ddot{\mathbf{r}}\mathbf{n}$) is (see the definition of relative curvature in Lecture 2) nothing but relative curvature $k(\mathbf{t})$. Thus

$$k(\mathbf{t}) = \ddot{\mathbf{r}}\mathbf{n}. \tag{3}$$

Since

$$\begin{aligned}\ddot{\mathbf{r}} &= \dot{\mathbf{r}}_u \dot{u} + \mathbf{r}_u \ddot{u} + \dot{\mathbf{r}}_v \dot{v} + \mathbf{r}_v \ddot{v}\\ &= (\mathbf{r}_{uu}\dot{u} + \mathbf{r}_{uv}\dot{v})\,\dot{u} + (\mathbf{r}_{uv}\dot{u} + \mathbf{r}_{vv}\dot{v})\,\dot{v} + \mathbf{r}_u \ddot{u} + \mathbf{r}_v \ddot{v}\\ &= \mathbf{r}_{uu}\dot{u}^2 + 2\mathbf{r}_{uv}\dot{u}\dot{v} + \mathbf{r}_{vv}\dot{v}^2 + \mathbf{r}_u \ddot{u} + \mathbf{r}_v \ddot{v}\end{aligned}$$

and $\mathbf{r}_u\mathbf{n} = 0$, $\mathbf{r}_v\mathbf{n} = 0$, this proves that

$$k(\mathbf{t}) = L\dot{u}^2 + 2M\dot{u}\dot{v} + N\dot{v}^2,$$

where

$$L = \mathbf{r}_{uu}\mathbf{n}, \quad M = \mathbf{r}_{uv}\mathbf{n}, \quad N = \mathbf{r}_{vv}\mathbf{n}. \tag{4}$$

Thus function $\mathbf{t} \mapsto k(\mathbf{t})$ can be easily extended to all possible tangent vectors $d\mathbf{r} \neq 0$ assuming by definition that

$$k(d\mathbf{r}) = k\left(\frac{d\mathbf{r}}{|d\mathbf{r}|}\right).$$

Since $|d\mathbf{r}|^2 = d\mathbf{r}^2 = ds^2$, where

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2$$

(see formula (28), Lecture 3) and the coordinates $\dot{u}$ and $\dot{v}$ of the vector $\frac{d\mathbf{r}}{|d\mathbf{r}|} = \frac{d\mathbf{r}}{ds}$ are equal to $\frac{du}{ds}$ and $\frac{dv}{ds}$, we have

$$k\,(d\mathbf{r}) = \frac{L\,du^2 + 2M\,du\,dv + N\,dv^2}{E\,du^2 + 2F\,du\,dv + G\,dv^2}\,. \tag{5}$$

[Notice that the equations $\dot{u} = \frac{du}{ds}$ and $\dot{v} = \frac{dv}{ds}$ formally coincide with the equations, known from calculus, for derivatives as the ratios of the differentials. However, they have another intentional meaning in our case since *du* and *dv* are the coordinates of the tangent vector $d\mathbf{r}$ rather than differentials and *ds* is the length of the vector.]

Definition 1. The quadratic form

$$L\,du^2 + 2M\,du\,dv + N\,dv^2$$

is called the *second quadratic form* of a surface $\mathscr{X}$ and is denoted by II.

In this notation formula (5) can be written as follows:

$$k = \frac{\mathrm{II}}{\mathrm{I}}. \tag{6}$$

For the coefficients of form II, besides formulas (4) there are also formulas

$$L = -\mathbf{r}_u\mathbf{n}_u, \quad M = -\mathbf{r}_u\mathbf{n}_v = -\mathbf{r}_v\mathbf{n}_u, \quad N = -\mathbf{r}_v\mathbf{n}_v. \tag{7}$$

Indeed, since $\mathbf{r}_u\mathbf{n} = 0$, we have $\mathbf{r}_{uu}\mathbf{n} + \mathbf{r}_u\mathbf{n}_u = 0$ and $\mathbf{r}_{uv}\mathbf{n} + \mathbf{r}_u\mathbf{n}_v = 0$, i.e. $L = -\mathbf{r}_u\mathbf{n}_u$ and $M = -\mathbf{r}_u\mathbf{n}_v$. Similarly, since $\mathbf{r}_v\mathbf{n} = 0$ we have $\mathbf{r}_{vu}\mathbf{n} + \mathbf{r}_v\mathbf{n}_u = 0$ and $\mathbf{r}_{vv}\mathbf{n} + \mathbf{r}_v\mathbf{n}_v = 0$, i.e. $M = -\mathbf{r}_v\mathbf{n}_u$ and $N = -\mathbf{r}_v\mathbf{n}_v$. ☐

Introducing the vector

$$d\mathbf{n} = \mathbf{n}_u du + \mathbf{n}_v dv \tag{8}$$

we can write form II (by virtue of (7)) as

$$\mathrm{II} = -d\mathbf{r}\,d\mathbf{n},$$

in a way similar to $\mathrm{I} = d\mathbf{r}^2$ for form I.

Remark 1. By analogy the *third quadratic form* $\mathrm{III} = d\mathbf{n}^2$ could also be introduced. But, as is to be shown below, it can be linearly expressed in terms of forms I and II and therefore gives nothing new. The coefficients *L*, *M*, *N* of form II are also denoted by *D*, *D′*, *D″*.

To represent the function $\mathbf{t} \mapsto k(\mathbf{t})$ graphically the French mathematician Dupin suggested considering on a tangent plane the curve (now called the *Dupin indicatrix*) that results when for any unit tangent vector $\mathbf{t}$ a segment of length $|k(\mathbf{t})|^{-1/2}$ is laid off from the point of tangency p (taken as the origin O on the tangent plane) in the direction of that vector. Denote by x and y the coordinates (in the coordinate system $O\mathbf{r}_u\mathbf{r}_v$) of the end point of the segment; then its length is expressed as follows:

$$|x\mathbf{r}_u + y\mathbf{r}_v| = \sqrt{\mathrm{I}(x, y)}.$$

Since the curvature $k(\mathbf{t})$ is expressed by formula (6), which in the present notation has the form

$$k(\mathbf{t}) = \frac{\mathrm{II}(x, y)}{\mathrm{I}(x, y)},$$

we obtain for the Dupin indicatrix the equation

$$\sqrt{\mathrm{I}(x, y)} = \sqrt{\frac{\mathrm{I}(x, y)}{\mathrm{II}(x, y)}},$$

i.e. the equation

$$|\mathrm{II}(x, y)| = 1.$$

This proves that the *Dupin indicatrix is a curve with equation*

$$|Lx^2 + 2Mxy + Ny^2| = 1.$$

When $LN - M^2 > 0$ the curve (more precisely, the set of its real points) is an ellipse with equation

$$Lx^2 + 2Mxy + Ny^2 = \varepsilon, \tag{9}$$

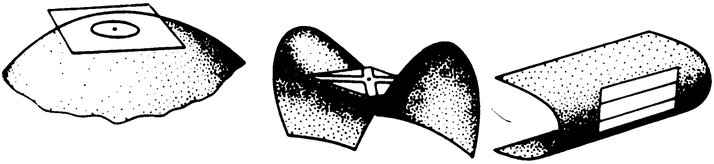
where $\varepsilon = +1$ if $L > 0$ and $\varepsilon = -1$ if $L < 0$. Accordingly a point of the surface at which $LN - M^2 > 0$ is called *elliptic*.

At an elliptic point all curvatures $k(\mathbf{t})$ have the same sign (coinciding with that of L). Among them, there is one maximum k_1 and one minimum k_2 curvatures (unless they all coincide, i.e. unless the Dupin indicatrix is a circle) corresponding to the directions of the minor and major axes of ellipse (9).

When $LN - M^2 < 0$ the Dupin indicatrix consists of two hyperbolas

$$Lx^2 + 2Mxy + Ny^2 = \pm 1 \tag{10}$$

with common asymptotes and therefore a point of the surface at which $LN - M^2 < 0$ is called *hyperbolic*. In the direction of the real axis of one of the hyperbolas (10) the curvature k (**t**) attains its maximum value $k_1 > 0$. As the vector **t** is rotated the curvature first decreases to zero when the vector **t** assumes the asymptotic direction, and then while continuing to decrease attains its minimum value $k_2 < 0$ when the direction of **t** coincides with that of the real axis of the other hyperbola (i.e. with the direction of the imaginary axis of the first hyperbola).


(1) (2) (3)

The Dupin indicatrix
(1) At an elliptic point
(2) At a hyperbolic point
(3) At a parabolic point

When $LN - M^2 = 0$ a point of the surface is called *parabolic*. At such a point the Dupin indicatrix has the equation

$$(\sqrt{|L|}\, x + \sqrt{|N|}\, y)^2 = 1 \tag{11}$$

and therefore is a pair of parallel lines (provided $L \neq 0$ or $N \neq 0$). In the direction of these lines the curvature k (**t**) is equal to zero, in the perpendicular direction it reaches its maximum value (in magnitude) while maintaining throughout the same sign. But if $L = 0$, $N = 0$ (and therefore $M = 0$), the curvature k (**t**) is identically as a function of **t** equal to zero (and the Dupin indicatrix is not defined).

Notice that at elliptic and parabolic points the Dupin indicatrix is a second degree curve, and at hyperbolic points it is a quartic curve.

In each of the three cases the function $k(\mathbf{t})$ attains *twice* its maximum k_1 and minimum k_2 (unless it is identically zero).

Problem 1. Prove that

$$k(\mathbf{t}) = k_1 \cos^2 \varphi + k_2 \sin^2 \varphi,$$

where φ is the angle made by the vector $\mathbf{t}$ and the direction in which the curvature is equal to k_1. This formula is known as the *Euler formula.*

Definition 2. Numbers k_1 and k_2 are called the *principal curvatures* of the surface at the point under consideration. Their product

$$K = k_1 k_2$$

is called the *total* (or *Gaussian*) *curvature* and their half-sum

$$H = \frac{k_1 + k_2}{2}$$

is called *mean curvature.*

According to the foregoing, $K > 0$ at an elliptical point, $K < 0$ at a hyperbolic point, and $K = 0$ at a parabolic point.

We stress that H and K depend on the point $p \in \mathscr{X}$, i.e. are functions on $\mathscr{X}$. As functions of local coordinates these functions are smooth.

To find principal curvatures one could seek for the principal directions of the second degree curves (9) and (10) (there is no problem with curve (11)) and then find their canonical equations. This method, however, involves lengthy computations because the coordinates x and y are not rectangular. Therefore we shall proceed in a different way, directly applying the basic formula (6).

According to this formula curvature k_2 is the smallest value of the function

$$\frac{\mathrm{II}(x, y)}{\mathrm{I}(x, y)} = \frac{Lx^2 + 2Mxy + Ny^2}{Ex^2 + 2Fxy + Gy^2}$$

of two variables x and y, with $(x, y) \neq (0, 0)$. Hence

$$\frac{\mathrm{II}(x, y)}{\mathrm{I}(x, y)} \geqslant k_2$$

for all $(x, y) \neq (0, 0)$, the equality holding at least at one point (x, y). Since $\mathrm{I}(x, y) > 0$ when $(x, y) \neq (0, 0)$, this inequality is equivalent to the inequality

$$\mathrm{II}(x, y) - k_2\mathrm{I}(x, y) \geqslant 0$$

which means that the quadratic form $\mathrm{II} - k_2\mathrm{I}$ with matrix

$$\left\| \begin{matrix} L-k_2E & M-k_2F \\ M-k_2F & N-k_2G \end{matrix} \right\|$$

is nonnegative at all points $(x, y) \neq (0, 0)$ and zero at least at one of them.

Similarly, the number k_1 is characterized by the fact that the quadratic form $\mathrm{II} - k_1\mathrm{I}$ is everywhere positive and zero at least at one point $(x, y) \neq (0, 0)$.

But it is easy to see (directly or on the basis of the general theory of quadratic forms over the field $\mathbb{R}$; see Lecture II.12) that a *quadratic form in two variables is everywhere nonpositive or nonnegative and zero at least at one point* $(x, y) \neq (0, 0)$ *if and only if its rank is less than two, i.e. if the determinant of its matrix is zero.*

This proves that the *principal curvatures* k_1, k_2 *are the roots of the equation*

$$\begin{vmatrix} L-kE & M-kF \\ M-kF & N-kG \end{vmatrix} = 0,$$

i.e. of the equation

$$(EG - F^2)k^2 - (EN + GL - 2FM)k + (LN - M^2) = 0.$$

In particular, it follows (by virtue of Viète's formulas) that

$$K = \frac{LN-M^2}{EG-F^2}, \quad H = \frac{1}{2}\frac{EN+GL-2FM}{EG-F^2}.$$

The first of these formulas will find an important application in our next lecture.

Suppose that the coordinates x, y, and z in a space $\mathcal{A}$ have been chosen so that the surface under consideration is the graph of a function $z = z(x, y)$, with $z(0, 0) = 0$ and the normal vector at the point $(0, 0)$ being the unit vector $\mathbf{k}$ of the Oz-axis (see above).

Then x and y are, in a neighbourhood of the point $(0, 0)$, the coordinates u and v on the surface, with

$$\mathbf{r}_u = (1, 0, z_x), \quad \mathbf{r}_v = (0, 1, z_y).$$

Since at $(0, 0)$ the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ are parallel by the condition for the coordinate plane Oxy, it follows that

$$(z_x)_0 = 0 \quad \text{and} \quad (z_y)_0 = 0$$

(with the subscript 0 we write the values of partial derivatives at $(0, 0)$) and hence the expansion of the function $z(x, y)$ into a Taylor series begins with quadratic terms

$$z = \frac{1}{2}(rx^2 + 2sxy + ty^2) + \ldots,$$

where

$$r = (z_{xx})_0, \quad s = (z_{xy})_0, \quad t = (z_{yy})_0$$

(surface of Monge).

On the other hand, since

$$\mathbf{r}_{uu} = (0, 0, z_{xx}), \ \mathbf{r}_{uv} = (0, 0, z_{xy}), \ \mathbf{r}_{vv} = (0, 0, z_{yy}),$$

at $(0, 0)$ we have

$$L = r, \quad M = s, \quad N = t.$$

Thus, in this case, the *second quadratic form of a surface differs only by a constant factor of* $\frac{1}{2}$ *from the sum* $z_2(x, y)$ *of the quadratic terms in the Taylor series of the function* $z(x, y)$.

Since near the point $(0, 0)$ a surface $z = z(x, y)$ differs but little from a surface $z = z_2(x, y)$ and since for $rt - s^2 > 0$ the latter surface is an elliptic paraboloid and for $rt - s^2 < 0$ it is a hyperbolic paraboloid, this proves that an *arbitrary surface differs but little from an elliptic paraboloid near the elliptic point and from a hyperbolic paraboloid near the hyperbolic point.*

This gives quite a satisfactory idea of the behaviour of a surface near nonparabolic points.

When $rt - s^2 = 0$ but either $r \neq 0$ or $s \neq 0$, the surface $z = z_2(x, y)$ is a parabolic cylinder. *About a parabolic point for which either $L \neq 0$ or $N \neq 0$ the surface therefore differs but little from the parabolic cylinder.*

As to the behaviour of a surface near a parabolic point at which $L = 0$ and $N = 0$ (and hence $M = 0$) nothing definite can be said; it may generally be quite complex.

The computations also show that at (0, 0) we have $E = 1$, $F = 0$ and $G = 0$ for the surface under consideration, from which it follows that at this point

$$K = rt - s^2 \quad \text{and} \quad H = \frac{r+t}{2}.$$

Moreover, immediate calculations (which can be done mentally if we notice that for functions f and g having the property that $f(0) = 0$ and $g(0) = 1$ the derivative $(fg)'$ of their product fg assumes at zero the value $f'(0)$) shows that at (0, 0)

$$\mathbf{n}_u = (-r, -s, 0), \quad \mathbf{n}_v = (-s, -t, 0)$$

and hence

$$\mathbf{n}_u^2 = r^2 + s^2, \quad \mathbf{n}_u\mathbf{n}_v = s(r + t), \quad \mathbf{n}_v^2 = s^2 + t^2.$$

It follows that

$$\mathbf{n}_u^2 = 2HL - KE, \quad \mathbf{n}_u\mathbf{n}_v = 2HM - KF, \quad \mathbf{n}_v^2 = 2HN - KG,$$

i.e. that

$$\mathrm{III} = 2H\mathrm{II} - K\mathrm{I},$$

where III is the third quadratic form of a surface introduced in Remark 1. Thus form III is indeed linearly expressed in terms of forms I and II.

For the ruled surface

$$\mathbf{r} = \boldsymbol{\rho}(u) + v\mathbf{a}(u), \tag{12}$$

as we already know,

$$E = 1 + 2v\dot{\boldsymbol{\rho}}\mathbf{a} + v^2\dot{\mathbf{a}}^2, \quad F = \dot{\boldsymbol{\rho}}\mathbf{a}, \quad G = 1$$

(we assume as usual that the parameter u on the curve $\boldsymbol{\rho} = \boldsymbol{\rho}(u)$ is natural and the vector $\mathbf{a}(u)$ is a unit vector). Further

$$\mathbf{r}_u = \dot{\boldsymbol{\rho}} + v\dot{\mathbf{a}}, \quad \mathbf{r}_v = \mathbf{a},$$

$$\mathbf{r}_u \times \mathbf{r}_v = \dot{\boldsymbol{\rho}} \times \mathbf{a} + v(\dot{\mathbf{a}} \times \mathbf{a}),$$

$$\mathbf{n} = \frac{\dot{\boldsymbol{\rho}} \times \mathbf{a} + v(\dot{\mathbf{a}} \times \mathbf{a})}{\sqrt{EG-F^2}},$$

$$\mathbf{r}_{uu} = \ddot{\boldsymbol{\rho}} + v\ddot{\mathbf{a}}, \quad \mathbf{r}_{uv} = \dot{\mathbf{a}}, \quad \mathbf{r}_{vv} = 0,$$

$$L = \frac{(\ddot{\boldsymbol{\rho}} + v\ddot{\mathbf{a}})(\dot{\boldsymbol{\rho}} \times \mathbf{a} + v(\dot{\mathbf{a}} \times \mathbf{a}))}{\sqrt{EG-F^2}}, \quad M = \frac{\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}}}{\sqrt{EG-F^2}}, \quad N = 0,$$

$$LN - M^2 = -\frac{(\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}})^2}{EG-F^2},$$

and therefore

$$K = -\frac{(\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}})^2}{(EG-F^2)^2} \leqslant 0.$$

Thus the total curvature of an arbitrary ruled surface is nonpositive at any of its points, i.e. a *ruled surface has no elliptic points*.

When the surface is a cylinder ($\dot{\mathbf{a}} = 0$), a cone ($\mathbf{a} = \boldsymbol{\rho}$ and therefore $\dot{\mathbf{a}} = \dot{\boldsymbol{\rho}}$) or a surface of tangents ($\mathbf{a} = \dot{\boldsymbol{\rho}}$), the formula obtained yields $K = 0$. Thus the *total curvature of every developable is zero* (at any point).

Conversely, if $K = 0$, then $\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}} = 0$, i.e. the vectors $\dot{\boldsymbol{\rho}}, \mathbf{a}, \dot{\mathbf{a}}$ are coplanar. If the vector $\dot{\mathbf{a}}(u)$ is not identically zero, i.e. if surface (12) is not a cylinder, then, going over if necessary, to a smaller neighbourhood, we may assume that $\dot{\mathbf{a}}(u) \neq 0$ for all u. The vectors $\mathbf{a}$ and $\dot{\mathbf{a}}$ are therefore linearly independent (they are nonzero and orthogonal) and hence the vector $\dot{\boldsymbol{\rho}}$ is expressed linearly in terms of $\mathbf{a}$ and $\dot{\mathbf{a}}$:

$$\dot{\boldsymbol{\rho}} = \lambda\mathbf{a} + \mu\dot{\mathbf{a}},$$

where $\lambda = \lambda(u)$, $\mu = \mu(u)$ are some functions of u.

Let

$$u_1 = u, \quad v_1 = v + \mu(u).$$

Since the Jacobian of this transformation is equal to 1, the numbers u_1 and v_1 are also, after changing to a smaller neighbourhood, the coordinates on surface (12). In these coordinates the equation of the surface is of the form

$$\mathbf{r} = \boldsymbol{\rho}_1(u) + v\mathbf{a}(u)$$

(we omit the subscripts in u_1 and v_1), where

$$\boldsymbol{\rho}_1(u) = \boldsymbol{\rho}(u) - \mu(u)\,\mathbf{a}(u),$$

and so,

$$\dot{\boldsymbol{\rho}}_1 = \dot{\boldsymbol{\rho}} - \dot{\mu}\mathbf{a} - \mu\dot{\mathbf{a}} = (\lambda - \dot{\mu})\,\mathbf{a}.$$

If $\dot{\boldsymbol{\rho}}_1$ is identically zero (i.e. $\lambda = \dot{\mu}$), then the equation of the surface is of the form

$$\mathbf{r} = \text{const} + v\mathbf{a}(u)$$

and therefore the surface is a cone. Otherwise we may assume, diminishing if necessary the neighbourhood, that $\dot{\boldsymbol{\rho}}_1(u) \neq 0$ for all u. Changing to the natural parameter (and changing if necessary the sign of v) we see that $\dot{\boldsymbol{\rho}}_1 = \mathbf{a}$, i.e. that the surface under consideration is a surface of tangents.

Thus we have proved the following proposition.

Proposition 1. *A ruled surface has zero total curvature at every point,*

$$K = 0,$$

if and only if it is a developable. □

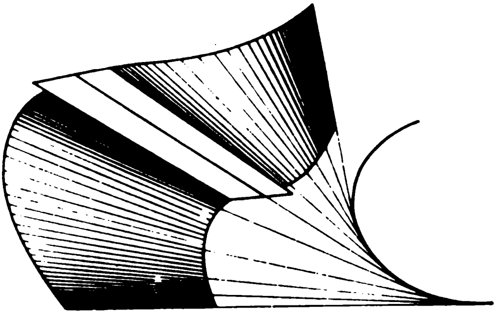
We have also established that developables are characterized by the condition

$$\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}} = 0$$

which is, as is easily seen, equivalent to the collinearity of the vectors $\dot{\boldsymbol{\rho}} \times \mathbf{a}$ and $\dot{\mathbf{a}} \times \mathbf{a}$. But the collinearity of these vectors implies that the vector

$$\mathbf{r}_u \times \mathbf{r}_v = \dot{\boldsymbol{\rho}} \times \mathbf{a} + v\,(\dot{\mathbf{a}} \times \mathbf{a})$$

is, up to proportionality, independent of v, i.e. the corresponding unit vector $\mathbf{n}$ is independent of v. This proves that *developables can be distinguished among all the ruled surfaces by the property that at all the points of each rectilinear generator such a surface has the same tangent plane.*



A developable of tangents

For an arbitrary surface of revolution

$$\mathbf{r} = x(v)\cos u\cdot\mathbf{i} + x(v)\sin u\cdot\mathbf{j} + z(v)\cdot\mathbf{k}$$

we have

$$\mathbf{r}_u = -x(v)\sin u\cdot\mathbf{i} + x(v)\cos u\cdot\mathbf{j},$$

$$\mathbf{r}_v = x'(v)\cos u\cdot\mathbf{i} + x'(v)\sin u\cdot\mathbf{j} + z'(v)\,\mathbf{k}$$

and hence $E = x(v)^2$, $F = 0$, $G = 1$ (we assume that $x'(v)^2 + z'(v)^2 = 1$; see Lecture 3). Therefore

$$\mathbf{r}_u \times \mathbf{r}_v = x(v)\,z'(v)\cos u\cdot\mathbf{i} + x(v)\,z'(v)\sin u\cdot\mathbf{j}$$
$$-x(v)\,x'(v)\,\mathbf{k},$$

$$\mathbf{n} = z'(v)\cos u\cdot\mathbf{i} + z'(v)\sin u\cdot\mathbf{j} - x'(v)\,\mathbf{k},$$

$$\mathbf{r}_{uu} = -x(v)\cos u\cdot\mathbf{i} - x(v)\sin u\cdot\mathbf{j},$$

$$\mathbf{r}_{uv} = -x'(v)\sin u\cdot\mathbf{i} + x'(v)\cos u\cdot\mathbf{j},$$

$$\mathbf{r}_{vv} = x''(v)\cos u\cdot\mathbf{i} + x''(v)\sin u\cdot\mathbf{j} + z''(v)\,\mathbf{k},$$

$$L = \mathbf{r}_{uu}\mathbf{n} = -x(v)\,z'(v), \quad M = \mathbf{r}_{uv}\mathbf{n} = 0,$$

$$N = \mathbf{r}_{vv}\mathbf{n} = x''(v)\,z'(v) - z''(v)\,x'(v) = -\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix},$$

$$\frac{LN-M^2}{EG-F^2} = \frac{z'(v)}{x(v)} \begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix}.$$

This proves that for a surface of revolution

$$K = \frac{z'(v)}{x(v)} \begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix}.$$

Example 1. For a sphere of radius R we have

$$x(v) = R\cos\frac{v}{R}, \quad z(v) = R\sin\frac{v}{R}$$

and therefore

$$x'(v) = -\sin\frac{v}{R}, \quad z'(v) = \cos\frac{v}{R},$$

$$x'' = -\frac{1}{R}\cos\frac{v}{R}, \quad z''(v) = -\frac{1}{R}\sin\frac{v}{R},$$

$$K = \frac{z'(v)}{x(v)} \begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = \frac{1}{R^2}.$$

Thus the *total curvature of a sphere of radius R is constant and equal to $1/R^2$*.

The result is intuitively obvious.

The following example is more interesting.

Example 2. A surface of revolution with profile

$$x(v) = R\sin(v), \quad z(v) = R\left(\ln\tan\frac{v}{2} + \cos v\right),$$

$$0 < v < \frac{\pi}{2}$$

(it is the so-called *tractrix*) is termed a *pseudosphere* (R is a *pseudoradius*). For this surface

$$x'(v) = R\cos v, \; z'(v) = \frac{R}{\sin v} - R\sin v = R\frac{\cos^2 v}{\sin v}$$

and hence

$$x'(v)^2 + z'(v)^2 = R^2\cot^2 v.$$

Since $x'(v)^2 + z'(v)^2 \neq 1$, the above general formula cannot be immediately applied. First it is necessary to go over to the natural parameter of the profile.

We have

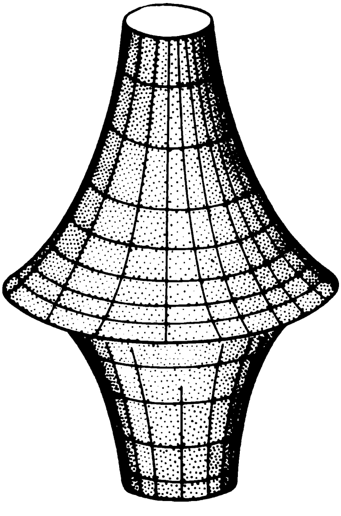
$$s = -R\int_{\frac{\pi}{2}}^{v} \cot v\,dv = -R\ln\sin v$$

and hence

$$\sin v = e^{-\frac{s}{R}}, \quad \cos v = \sqrt{1-e^{-2\frac{s}{R}}},$$

$$\tan\frac{v}{2} = e^{\frac{s}{R}} - \sqrt{e^{\frac{s}{R}}-1}.$$

Thus in terms of the natural parameter (which is again denoted by v) the tractrix will be given by the functions



A pseudosphere

$$x(v) = Re^{-\frac{v}{R}},$$

$$z(v) = R\ln\left(e^{\frac{v}{R}} - \sqrt{e^{2\frac{v}{R}}-1}\right) + R\sqrt{1-e^{-2\frac{v}{R}}}.$$

We calculate:

$$x'(v) = -e^{-\frac{v}{R}}, \quad z'(v) = -\sqrt{1-e^{-2\frac{v}{R}}},$$

$$x''(v) = \frac{1}{R}e^{-\frac{v}{R}}, \quad z''(v) = -\frac{e^{-2\frac{v}{R}}}{R\sqrt{1-e^{-2\frac{v}{R}}}},$$

$$\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = -\frac{e^{-\frac{v}{R}}}{R\sqrt{1-e^{-2\frac{v}{R}}}},$$

$$\frac{z'(v)}{x(v)}\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = -\frac{1}{R^2}.$$

Thus

$$K = -\frac{1}{R^2},$$

so that the *total curvature of a pseudosphere is constant and equal to* $-\frac{1}{R^2}$.

We see that in regard to total curvature the pseudosphere differs from a sphere only in the sense of curvature. This accounts for the term "pseudosphere".

Example 3. For the catenoid

$$x(v) = \cosh v, \quad z(v) = v,$$

$$x'(v) = \sinh v, \quad z'(v) = 1,$$

$$x'(v)^2 + z'(v)^2 = \cosh^2 v,$$

and therefore we must again pass to the natural parameter

$$s = \int_0^v \cosh v \, dv = \sinh v.$$

Again denoting this parameter by v we obtain the functions

$$x(v) = \sqrt{1+v^2}, \quad z(v) = \ln(v + \sqrt{1+v^2}).$$

Therefore

$$x'(v) = \frac{v}{\sqrt{1+v^2}}, \quad z'(v) = \frac{1}{\sqrt{1+v^2}},$$

$$x''(v) = \frac{1}{(1+v^2)^{3/2}}, \quad z''(v) = -\frac{v}{(1+v^2)^{3/2}},$$

$$\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = -\frac{1}{1+v^2},$$

and hence

$$K = -\frac{1}{(1+v^2)^2}.$$

It is interesting to compare the curvature of the catenoid with that of the isometric helicoid.

For the helicoid we have equation (12) with

$$\boldsymbol{\rho}(u) = u\cdot\mathbf{k} \text{ and } \mathbf{a}(u) = \cos u\cdot\mathbf{i} + \sin u\cdot\mathbf{j}.$$

Therefore

$$\dot{\boldsymbol{\rho}} = \mathbf{k}, \quad \dot{\mathbf{a}} = -\sin u\cdot\mathbf{i} + \cos u\cdot\mathbf{j},$$

$$E = 1 + 2v\dot{\boldsymbol{\rho}}\dot{\mathbf{a}} + v^2\dot{\mathbf{a}}^2 = 1 + v^2,$$

$$F = \dot{\boldsymbol{\rho}}\mathbf{a} = 0, \quad G = 1,$$

$$EG - F^2 = 1 + v^2,$$

$$\dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}} = \begin{vmatrix} 0 & 0 & 1 \\ \cos u & \sin u & 1 \\ -\sin u & \cos u & 0 \end{vmatrix} = 1,$$

and hence

$$K = -\frac{1}{(1+v^2)^2}.$$

We have obtained the same result as that for the catenoid! This means that *when the catenoid is bent into the helicoid the total curvatures at the corresponding points coincide.*

What happens to the mean curvature?

For the catenoid $E = 1 + v^2$, $F = 0$, $G = 1$. In addition

$$L = -x(v)\,z'(v) = -1,\quad M = 0,$$

$$N = -\begin{vmatrix} x'(v) & z'(v) \\ x''(v) & z''(v) \end{vmatrix} = \frac{1}{1+v^2},$$

and therefore

$$EN + GL - 2FM = 0,$$

i.e.

$$H = 0.$$

Thus *the mean curvature of the catenoid is equal to zero.*

For the helicoid

$$\ddot{\boldsymbol{\rho}} \times \mathbf{a} = \sin u \cdot \mathbf{i} - \cos u \cdot \mathbf{j},\quad \dot{\mathbf{a}} \times \mathbf{a} = -\mathbf{k},$$

$$\ddot{\boldsymbol{\rho}} = 0,\quad \ddot{\mathbf{a}} = -\cos u \cdot \mathbf{i} - \sin u \cdot \mathbf{j},$$

$$(\ddot{\boldsymbol{\rho}} + v\mathbf{a})\,(\ddot{\boldsymbol{\rho}} \times \mathbf{a} + v\,(\dot{\mathbf{a}} \times \mathbf{a})) = 0$$

and, in addition, as we have already seen,

$$E = 1 + v^2,\quad F = 0,\quad G = 1,$$

$$EG - F^2 = 1 + v^2,\quad \dot{\boldsymbol{\rho}}\mathbf{a}\dot{\mathbf{a}} = 0.$$

Therefore

$$L = 0,\quad M = \frac{1}{\sqrt{1+v^2}},\quad N = 0,$$

and hence

$$EN + LG - 2FM = 0,$$

i.e.

$$H = 0.$$

Thus the *mean curvature of the helicoid is also zero.*

The example of catenoid and helicoid suggests that total and mean curvatures are preserved under bending (isometry). It turns out that the hypothesis is true for the total curvature (we shall show this in our next lecture) whereas for the mean curvature it is false. Indeed, for a plane the mean curvature is zero while for a circular cylinder of radius R developable into a plane it is obviously equal to $1/2R$.

As to the reasons why the catenoid and helicoid have equal mean curvatures, we are deprived of the possibility of discussing them in this book.

Lecture 5

Weingarten formulas· Coefficients of connection· The Gauss theorem· Explicit formula for Gaussian curvature· The necessary and sufficient conditions of isometry· Surfaces of constant curvature

---

For the moving basis $\mathbf{r}_u$, $\mathbf{r}_v$, $\mathbf{n}$ of an arbitrary surface

$$\mathbf{r} = \mathbf{r}(u, v) \tag{1}$$

formulas can be written, similar to Frenet's formulas for curves, that yield an expansion of the derivatives

$$\mathbf{r}_{uu},\quad \mathbf{r}_{uv},\quad \mathbf{r}_{vv},\quad \mathbf{n}_u,\quad \mathbf{n}_v$$

of the vectors of the moving basis with respect to that same basis.

Since $\mathbf{n}^2 = 1$ and hence $\mathbf{n}\mathbf{n}_u = 0$ and $\mathbf{n}\mathbf{n}_v = 0$, the vectors $\mathbf{n}_u$ and $\mathbf{n}_v$ are only expanded with respect to the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$, so that

$$\mathbf{n}_u = \alpha\mathbf{r}_u + \beta\mathbf{r}_v,$$
$$\mathbf{n}_v = \alpha_1\mathbf{r}_u + \beta_1\mathbf{r}_v.$$

Multiplying the first of these formulas by $\mathbf{r}_u$ and $\mathbf{r}_v$ we obtain two relations:

$$-L = \mathbf{r}_u\mathbf{n}_u = \alpha\mathbf{r}_u^2 + \beta\mathbf{r}_u\mathbf{r}_v = \alpha E + \beta F,$$
$$-M = \mathbf{r}_v\mathbf{n}_u = \alpha\mathbf{r}_u\mathbf{r}_v + \beta\mathbf{r}_v^2 = \alpha F + \beta G,$$

from which it follows that

$$\alpha = \frac{FM - GL}{EG - F^2}, \quad \beta = \frac{FL - EM}{EG - F^2}.$$

The coefficients of the second formula are calculated in the same way:

$$\alpha_1 = \frac{FN - GM}{EG - F^2}, \quad \beta_1 = \frac{FM - EN}{EG - F^2}.$$

Further, since by definition

$$\mathbf{r}_{uu}\mathbf{n} = L, \quad \mathbf{r}_{uv}\mathbf{n} = M, \quad \mathbf{r}_{vv}\mathbf{n} = N$$

and since under the hypothesis $\mathbf{r}_u\mathbf{n} = 0$, $\mathbf{r}_v\mathbf{n} = 0$, the coefficients of $\mathbf{n}$ in the decomposition of vectors $\mathbf{r}_{uu}$, $\mathbf{r}_{uv}$, $\mathbf{r}_{vv}$ in terms of the basis $\mathbf{r}_u$, $\mathbf{r}_v$, $\mathbf{n}$ are equal to L, M, N respectively.

We thus have

$$(2) \qquad \begin{aligned} \mathbf{r}_{uu} &= \Gamma^1_{11}\mathbf{r}_u + \Gamma^2_{11}\mathbf{r}_v + L\mathbf{n}, \\ \mathbf{r}_{uv} &= \Gamma^1_{12}\mathbf{r}_u + \Gamma^2_{12}\mathbf{r}_v + M\mathbf{n}, \\ \mathbf{r}_{vv} &= \Gamma^1_{22}\mathbf{r}_u + \Gamma^2_{22}\mathbf{r}_v + N\mathbf{n}, \\ \mathbf{n}_u &= \frac{FM - GL}{EG - F^2}\mathbf{r}_u + \frac{FL - EM}{EG - F^2}\mathbf{r}_v, \\ \mathbf{n}_v &= \frac{FN - GM}{EG - F^2}\mathbf{r}_u + \frac{FM + EN}{EG - F^2}\mathbf{r}_v, \end{aligned}$$

where Γ^k_{ij}, $i, j, k = 1, 2$, are some functions of u and v. Formerly these functions were denoted by

$$\begin{Bmatrix} i \; j \\ k \end{Bmatrix}$$

and called the *Christoffel symbols*. But now they are usually called *connection coefficients*.

Formulas (2) are called *Weingarten formulas*.

To compute connection coefficients Γ^k_{ij} we first find the six products of vectors $\mathbf{r}_{uu}$, $\mathbf{r}_{uv}$, $\mathbf{r}_{vv}$ by vectors $\mathbf{r}_u$ and $\mathbf{r}_v$.

Since $\mathbf{r}_u^2 = E$, we have $2\mathbf{r}_{uu}\mathbf{r}_u = E_u$ and $2\mathbf{r}_{uv}\mathbf{r}_u = E_v$, i.e.

$$\mathbf{r}_{uu}\mathbf{r}_u = \frac{1}{2}E_u \text{ and } \mathbf{r}_{uu}\mathbf{r}_u = \frac{1}{2}E_v.$$

Similarly, since $\mathbf{r}_v^2 = G$ we have

$$\mathbf{r}_{uv}\mathbf{r}_v = \frac{1}{2}G_u \text{ and } \mathbf{r}_{vv}\mathbf{r}_v = \frac{1}{2}G_v.$$

Besides, since $\mathbf{r}_u\mathbf{r}_v = F$, we have $\mathbf{r}_{uu}\mathbf{r}_v + \mathbf{r}_u\mathbf{r}_{uv} = F_u$ and $\mathbf{r}_{uv}\mathbf{r}_v + \mathbf{r}_u\mathbf{r}_{vv} = F_v$, from which it follows that

$$\mathbf{r}_{uu}\mathbf{r}_v = F_u - \frac{1}{2}E_v \quad \text{and} \quad \mathbf{r}_{vv}\mathbf{r}_u = F_v - \frac{1}{2}G_u.$$

Now multiplying the first three of the formulas (2) by $\mathbf{r}_u$ and $\mathbf{r}_v$ we obtain six relations:

$$(3) \qquad \begin{cases} E\Gamma^1_{11} + F\Gamma^2_{11} = \frac{1}{2}E_u, \\ E\Gamma^1_{11} + G\Gamma^2_{11} = F_u - \frac{1}{2}E_v, \end{cases} \quad \begin{cases} E\Gamma^1_{12} + F\Gamma^2_{12} = \frac{1}{2}E_v, \\ E\Gamma^1_{12} + G\Gamma^2_{12} = \frac{1}{2}G_u, \end{cases} \quad \begin{cases} E\Gamma^1_{22} + F\Gamma^2_{22} = F_v - \frac{1}{2}G_u, \\ F\Gamma^1_{22} + G\Gamma^2_{22} = \frac{1}{2}G_v, \end{cases}$$

from which it is easy to find the coefficients Γ^k_{ij}. (The equations are uniquely solvable since the determinant $EG - F^2$ of each pair of equations is nonzero.)

We see that the *connection coefficients* Γ^k_{ij} *can be expressed in terms of the coefficients of the first quadratic form and of their derivatives.* Hence *they remain unaltered under bendings* (*isometries*) *of a surface.*

We shall not need explicit expressions for the coefficients Γ^k_{ij} in terms of the coefficients of the first quadratic form, and so we shall not give them here.

The coefficients of derived equations are connected by three relations resulting from calculating partial derivatives $\mathbf{r}_{uuv}$, $\mathbf{r}_{uvv}$, and $\mathbf{n}_{uv}$ in two different ways by using these formulas. One of these relations was found by Gauss and the other two by Peterson, Mainardi, and Codazzi. We shall consider only Gauss' relation which we shall obtain by calculating the coefficient of $\mathbf{r}_v$ in the expansion of the partial derivative $\mathbf{r}_{uuv}$ in terms of the vectors $\mathbf{r}_u$, $\mathbf{r}_v$, and $\mathbf{n}$.

In this calculation we shall only follow the coefficient of $\mathbf{r}_v$ and only those of its terms which depend on the coefficients of the second quadratic form. All the other terms will be replaced by dots.

We have

$$\begin{aligned}
\mathbf{r}_{uuv} &= (\mathbf{r}_{uu})_v = (\Gamma^1_{11}\mathbf{r}_u + \Gamma^2_{11}\mathbf{r}_v + L\mathbf{n})_v \\
&= \ldots + \Gamma^1_{11}\mathbf{r}_{uv} + \ldots + \Gamma^2_{11}\mathbf{r}_{vv} + \ldots + L\mathbf{n}_v \\
&= \ldots + \Gamma^1_{11}(\ldots) + \ldots + \Gamma^2_{11}(\ldots) \\
&\quad + \cdots + L\left(\cdots + \frac{FM-EN}{EG-F^2}\,\mathbf{r}_v\right) \\
&= \left(L\,\frac{FM-EN}{FG-F^2} + \cdots\right)\mathbf{r}_v + \ldots
\end{aligned}$$

Similarly

$$\begin{aligned}
\mathbf{r}_{uuv} &= (\mathbf{r}_{uv})_u = (\Gamma^1_{12}\mathbf{r}_u + \Gamma^2_{22}\mathbf{r}_v + M\mathbf{n})_u \\
&= \left(M\,\frac{FL-EM}{EG-F^2} + \cdots\right)\mathbf{r}_v + \ldots
\end{aligned}$$

Hence

$$L\,\frac{FM-EN}{EG-F^2} = M\,\frac{FL-EM}{EG-F^2} + \cdots,$$

where dots denote terms dependent only on the coefficients of the first quadratic form. But

$$M\,\frac{FL-EM}{EG-F^2} - L\,\frac{FM-EN}{EG-F^2} = E\,\frac{LN-M^2}{EG-F^2} = EK.$$

Since $E \neq 0$ (form I is positive definite), this proves that the *total curvature K of a surface is expressed in terms of the coefficients of the first quadratic form* (*and of their derivatives*). It follows that the curvature K remains unchanged under bendings. More precisely, if f is an isometry of a surface $\mathcal{X}$ onto a surface $\mathcal{Y}$, then

$$K_{\mathcal{Y}} \circ f = K_{\mathcal{X}}, \tag{4}$$

where $K_{\mathcal{X}}$ and $K_{\mathcal{Y}}$ are the total curvatures on $\mathcal{X}$ and $\mathcal{Y}$ respectively. (Indeed, if f preserves the coordinates of vectors in another system of coordinates, then both sides in (4) differ only in the notation of the coordinates.) This result deserves to be distinguished as a theorem.

Theorem 1 (Gauss' theorem). *The total (Gaussian) curvature of a surface remains unchanged under bendings (isometries), i.e. isometric surfaces have the same curvature at corresponding points.* □

Gauss was so delighted with the theorem that he called it the theorema egregium, which means a "brilliant theorem" in Latin.

From Theorem 1 it follows in particular that *no arbitrarily small part of a sphere can be bent into a plane.* Therefore no map gives an accurate representation of the Earth's surface.

An explicit expression of curvature K in terms of the coefficients E, F, and G of the first quadratic form is

$$(5)\qquad K = -\frac{1}{4(EG-F^2)^2}\begin{vmatrix} E & E_u & E_v \\ F & F_u & F_v \\ G & G_u & G_v \end{vmatrix} - \frac{1}{2\sqrt{EG-F^2}}\left\{\left(\frac{E_v-F_u}{\sqrt{EG-F^2}}\right)_v - \left(\frac{F_v-G_u}{\sqrt{EG-F^2}}\right)_u\right\}.$$

The other two relations resulting from differentiating the derived equations (and usually called the *Peterson-Codazzi formulas*) are of the form

$$2(EG-F^2)(L_v-M_u) - (EN+GL-2FM)(E_v-F_u) + \begin{vmatrix} E & E_u & L \\ F & F_u & M \\ G & G_u & N \end{vmatrix} = 0,$$

$$(6)\quad 2(EG-F^2)(M_v-N_u) - (EN+GL-2FM)(F_v-G_u) + \begin{vmatrix} E & E_v & L \\ F & F_v & M \\ G & G_v & N \end{vmatrix} = 0.$$

To prove these formulas all one needs is patience and care.

We shall prove only formula (5) and only for the case where $E = 1$ and $F = 0$, i.e. where the first quadratic form of a surface is expressed as

(7) $$\mathrm{I} = du^2 + G\,dv^2.$$

In this case equations (3) for Γ^k_{ij} are of the form

$$\Gamma^1_{11} = 0, \qquad \Gamma^1_{12} = 0, \qquad \Gamma^1_{22} = -\frac{1}{2} G_u,$$

$$G\Gamma^2_{11} = 0, \quad G\Gamma^2_{12} = G_u, \quad G\Gamma^2_{22} = \frac{1}{2} G_v,$$

from which it follows that

$$\Gamma^1_{11} = 0, \quad \Gamma^1_{12} = 0, \quad \Gamma^2_{11} = 0,$$

$$\Gamma^1_{22} = -\frac{1}{2} G_u, \quad \Gamma^2_{12} = \frac{1}{2} \frac{G_u}{G}, \quad \Gamma^2_{22} = \frac{1}{2} \frac{G_v}{G}.$$

Hence

$$\mathbf{r}_{uu} = L\mathbf{n} \quad \text{and} \quad \mathbf{r}_{uv} = \frac{1}{2} \frac{G_u}{G} \mathbf{r}_v + M\mathbf{n}.$$

Since in the case at hand

$$\mathbf{n}_u = -L\mathbf{r}_v - \frac{M}{G} \mathbf{r}_v \quad \text{and} \quad \mathbf{n}_v = -M\mathbf{r}_u - \frac{N}{G} \mathbf{r}_v,$$

it follows that

$$\mathbf{r}_{uuv} = (L\mathbf{n})_v = L_v\mathbf{n} + L\left(-M\mathbf{r}_u - \frac{N}{G}\mathbf{r}_v\right)$$
$$= -LM\mathbf{r}_u - \frac{2N}{G}\mathbf{r}_v + L_v\mathbf{n}$$

and

$$\mathbf{r}_{uvu} = \left(\frac{1}{2}\frac{G_u}{G}\mathbf{r}_v + M\mathbf{n}\right)_u = \frac{1}{2}\left(\frac{G_u}{G}\right)_u \mathbf{r}_v$$
$$+ \frac{1}{2}\frac{G_u}{G}\left[\frac{1}{2}\frac{G_u}{G}\mathbf{r}_v + M\mathbf{n}\right]$$
$$+ M\mathbf{n} + M\left(-L\mathbf{r}_u - \frac{M}{G}\mathbf{r}_v\right)$$
$$= -LM\mathbf{r}_u + \left[\frac{1}{2}\left(\frac{G_u}{G}\right)_u + \frac{1}{4}\left(\frac{G_u}{G}\right)^2 - \frac{M^2}{G}\right]\mathbf{r}_v$$
$$+ \left(\frac{1}{2}\frac{G_u}{G}M + M_u\right)\mathbf{n}$$

and hence

$$-\frac{LN}{G}=\frac{1}{2}\left(\frac{G_u}{G}\right)+\frac{1}{4}\left(\frac{G_u}{G}\right)^2-\frac{M^2}{G},$$

$$L_v=\frac{1}{2}\frac{G_u}{G}M+M_u.$$

The second equation is now of no interest to us (it is the first of equations (6) for $E=1$ and $F=0$) and the first (since in the case at hand $K=\frac{1}{G}(LN-M^2)$) gives

$$K=-\frac{1}{2}\left(\frac{G_u}{G}\right)_u-\frac{1}{4}\left(\frac{G_u}{G}\right)^2,$$

i.e., as an obvious computation shows,

$$K=-\frac{(\sqrt{G})_{uu}}{\sqrt{G}}, \tag{8}$$

which coincides with the result of substituting the values of $E=1$ and $F=0$ in formula (5).

Thus we have proved that the *total curvature of a surface with the first quadratic form* (7) *is expressed by formula* (8).

Let, for example,

$$\mathrm{I}=du^2+\cos^2 u\ dv^2. \tag{9}$$

Then $\sqrt{G}=\cos u$ and $(\sqrt{G})_{uu}=-\cos u$. Hence $K=1$, which agrees quite well with the result of Example 1 of Lecture 4 (for (9) is the first quadratic form of a sphere of radius $R=1$; see formula (36) of Lecture 3, in which u and v are interchanged).

Similarly, it can be shown that a surface with the first quadratic form

$$\mathrm{I}=du^2+\cosh^2 u\ dv^2$$

has the curvature $K=-1$ (cf. Example 2 of Lecture 4).

Remark 1. We stress that all these results hold for surfaces in a *three-dimensional* Euclidean space. For surfaces in a larger, three-dimensional space, however, (5) (or its special form (8) may be taken to be a definition of curvature K.

Remark 2. For six functions

$$E,\ F,\ G,\ L,\ M,\ N \tag{10}$$

given in an open convex set $U \subset \mathbb{R}^2$ to be the coefficients of the first and the second quadratic form of some surface $\mathbf{r} = \mathbf{r}(u, v)$ it is necessary that for these functions, in addition to conditions

$$E > 0, \quad EG - F^2 > 0 \tag{11}$$

of positive definiteness, relations (5) and (6) hold (it is meant that we have substituted $K = \frac{LM - M^2}{EG - F^2}$ in relation (5)). It turns out that *these relations are also sufficient* (for the existence of a regular, but not generally elementary, surface with the given forms I and II). Moreover, functions (10) (satisfying relations (5), (6), and (11)) *define a surface up to movement of space.* These statements are the two-dimensional analogue of the corresponding statements for curves (see Theorem 1 of Lecture 2) and can be proved by a similar method (but instead of the unique existence theorem for solutions of linear ordinary differential equations we use the corresponding theorem for the system of linear partial differential equations).

The Gauss theorem states that for the two surfaces to be isometric it is necessary that the total curvatures be equal. Although this condition is by no means sufficient, it is so strong that making use of it sufficient conditions can be easily obtained. We shall not expound this question and only consider the most important special case of the corresponding theorem.

Let

$$\Delta_1 K = \frac{EK_v^2 - 2FK_uK_v + GK_u^2}{EG - F^2}$$

be Beltrami's first differential parameter of the function K. If the two functions K and $\Delta_1 K$ of u and v are *functionally independent*, i.e. if their Jacobian

$$\begin{vmatrix} \dfrac{\partial K}{\partial u} & \dfrac{\partial K}{\partial v} \\ \dfrac{\partial \Delta_1 K}{\partial u} & \dfrac{\partial \Delta_1 K}{\partial v} \end{vmatrix}$$

is everywhere nonzero, then they may be taken as new local coordinates on the surface. We call them *Gaussian coordinates*. From the property that the operator Δ_1 is invariant (formula (34) of Lecture 3) and formula (4) it follows directly that for any isometry $f\colon \mathcal{X} \to \mathcal{Y}$ the equality

$$\Delta_1 K_{\mathcal{Y}} \circ f = \Delta_2 K_{\mathcal{X}} \tag{12}$$

holds. Taken together, formulas (4) and (12) imply that *every isometry is a mapping which preserves the coordinates of vectors in another system of coordinates.* The following theorem is therefore true.

Theorem 2. *Two elementary surfaces on which Gaussian coordinates are defined are isometric if and only if the first quadratic forms in these coordinates coincide.* ☐

Thus, to determine whether or not two surfaces are isometric it is necessary to introduce (if possible) Gaussian coordinates and calculate in these coordinates the first quadratic forms of the surfaces. If the forms coincide, the surfaces are isometric, but if they differ, the surfaces are not isometric.

Theorem 2 gives no answer when K and $\Delta_1 K$ are functionally dependent, for example, when $\Delta_1 K = 0$ (which occurs, as can be easily figured out, if and ony if $K =$ const). In this case, however, the condition of Theorem 1 proves to be sufficient, i.e. *two elementary surfaces of constant total curvature are isometric if and only if they have the same curvature.* In other words, *any surface of constant total curvature K is locally isometric to a sphere of radius $R = \frac{1}{\sqrt{K}}$ if $K > 0$, to a plane if $K = 0$, and to a pseudosphere of pseudoradius $R = \frac{1}{\sqrt{-K}}$ if $K < 0$.*

To prove this we shall need a lemma which is to be proved in the next semester:

Lemma (Gauss). *On any surface there are local coordinates u, v in which the first quadratic form of that surface is of form (7), the function $G = G(u, v)$ having the property that*

$$G(0, v) = 1 \quad \textit{and} \quad G_u(0, v) = 0 \textit{ for all } v. \tag{13}$$

By virtue of this lemma we may assume without loss of generality that the first quadratic form of a given surface of constant total curvature K is of form (7) and hence formula (8) holds for K. This formula may be thought of as a differential equation of second degree with constant coefficients relative to the function $\sqrt{G}$. It is known from the theory of differential equations that the general solution of that equation is of the form

$$\sqrt{G}=\begin{cases} A\cos a(u+B) & \text{if } K=a^2>0,\\ Au+B & \text{if } K=0,\\ A\cosh a(u+B) & \text{if } K=-a^2<0,\end{cases}$$

where A and B are arbitrary functions of v. But in view of the first of the conditions (13) we must have

$$\begin{aligned} A\cos aB&=1 \quad \text{for } K>0,\\ B&=1 \quad \text{for } K=0,\\ A\cosh aB&=1 \quad \text{for } K<0,\end{aligned}$$

and in view of the second of the conditions (13), by virtue of the identities

$$(\sqrt{G})_u=\frac{1}{2}\frac{G_u}{\sqrt{G}}=\begin{cases} -Aa\sin a(u+B) & \text{if } K>0,\\ A & \text{if } K=0,\\ Aa\sinh a(u+B) & \text{if } K<0\end{cases}$$

we must have

$$\begin{aligned} Aa\sin aB&=0 \quad \text{for } K>0,\\ A&=0 \quad \text{for } K=0,\\ Aa\sinh aB&=0 \quad \text{for } K<0.\end{aligned}$$

Hence

$$\sqrt{G}=\begin{cases} \cos au & \text{if } K=a^2>0,\\ 1 & \text{if } K=0,\\ \cosh au & \text{if } K=-a^2<0.\end{cases}$$

In the first case we obtain the first quadratic form

$$du^2+\cos^2 au\, dv^2$$

for a sphere of radius $1/a$, in the second case we obtain the first quadratic form

$$du^2 + dv^2$$

for a plane, and in the third the first quadratic form

$$du^2 + \cosh^2 au\, dv^2$$

for a pseudosphere of pseudoradius $1/a$. □

Here we interrupt the exposition of the theory of surfaces and turn to the main subject of this course, the theory of smooth manifolds.

Lecture 6

Introductory remarks · Open subsets of the space $\mathbb{R}^n$ and their diffeomorphisms · Charts and atlases · Maximal atlases · Smooth manifolds · Examples of smooth manifolds

---

The concept of a smooth manifold is one of the basic concepts of modern mathematics. It arises as a result of explicating and generalizing simultaneously to greater dimensions an intuitive idea of a surface considered irrespective of its position in space. Basic principles of explicating the idea are taken from cartography.

Individual regions of the Earth's surface can be adequately described by means of maps or charts which allow a region to be represented in a plane. Every point of the Earth can be represented on a map or a chart but one map is insufficient to cover the entire Earth; an atlas is required to do this, i.e. one needs a system of maps or charts. Any chart allows us to map rectilinear coordinates in the plane into an appropriate region on the surface and thus to obtain *local coordinates* in it. (On the contrary, in practical mathematical cartography geographical coordinates on the Earth's surface are mapped into a curvilinear coordinate network in the plane, but this difference is of no fundamental importance.) The local coordinates corresponding to two different charts are connected by transition functions which allow one to express some coordinates (in the common part of the two charts) in terms of the others.

In the corresponding general definitions we replace the plane by a standard Euclidean space $\mathbb{R}^n$, where n is some integer which is assumed to be chosen and fixed. The degenerate case of $n = 0$ is not excluded for the time being.

First of all recall from calculus some facts and definitions concerning the space $\mathbb{R}^n$ (which we have already used partly in the preceding lectures).

A point $\mathbf{x}$ of a subset $U \subset \mathbb{R}^n$ is said to be an *interior point* of the subset if there is an $\varepsilon > 0$ such that a ball of radius ε with centre at $\mathbf{x}$ is contained entirely in U. The set U is said to be *open* (in $\mathbb{R}^n$) if all its points are interior points.

An arbitrary mapping $\varphi: U \to V$ of subsets of $\mathbb{R}^n$ is given by n functions

(1) $$y^1 = \varphi^1(x^1, \ldots, x^n), \ldots, y^n = \varphi^n(x^1, \ldots, x^n)$$

of n variables which express, in terms of coordinates $x^1, \ldots, x^n$ of an arbitrary point $\mathbf{x} \in U$, the coordinates $y^1, \ldots, y^n$ of a point $\mathbf{y} = \varphi(\mathbf{x})$. When U is an open set, it makes sense to speak of the derivatives of arbitrary order of functions (1) with respect to $x^1, \ldots, x^n$ at any point of U.

A mapping φ of an open set U is said to be a *C^r-mapping*, where r is some natural number or a symbol ∞, if at every point of U functions (1) have continuous partial derivatives of all orders $\leqslant r$ for $r \neq \infty$ and continuous derivatives of all orders for $r = \infty$. But when at any point of U functions (1) are real analytic (are expanded into power series with a nonzero radius of convergence), the mapping φ is said to be a *C^ω-mapping*.

In what follows we shall assume some C^r-smoothness, where $r \leqslant \infty$ or $r = \omega$, to be fixed, and mappings of this class will be simply called *smooth mappings*. As a rule, any $r \geqslant 2$ and often even $r = 1$ will suit us, but in order not to have to watch whether derivatives of too high orders appear we shall assume, unless otherwise stated, that $r = \infty$ or $r = \omega$. We shall sometimes consider also the particular case $r = 0$, however. In this case smooth mappings are simply continuous mappings.

The case of $r = \infty$ will be assumed to be the main, and all situations where it differs from $r = \omega$ will be stated especially.

Every smooth mapping $\varphi: U \to V$ defines on U a smooth function

$$D\varphi = \det \left\| \frac{\partial \varphi^i}{\partial x^j} \right\|, \quad i, j = 1, \ldots, n$$

called the *Jacobian* (or Jacobian determinant) (and also is denoted by J_φ). As shown in calculus, for any smooth mappings $\varphi: U \to V$ and $\psi: V \to W$, where U, V, and W are open subsets of $\mathbb{R}^n$, the composite mapping $\psi \circ \varphi: U \to W$ (acting according to the formula $(\psi \circ \varphi)(\mathbf{x}) = \psi(\varphi(\mathbf{x}))$, $\mathbf{x} \in U$) is also smooth, and

(2) $$D(\psi \circ \varphi)(\mathbf{x}) = (D\psi)(\mathbf{y}) \cdot (D\varphi)(\mathbf{x}), \quad \mathbf{y} = \varphi(\mathbf{x}),$$

for any point $\mathbf{x} \in U$ (formula (2) is usually called the **chain rule**, and the mapping $\psi \circ \varphi$ is the *composition* of mappings φ and ψ).

A mapping φ: $U \to V$ of open sets of a space $\mathbb{R}^n$ is called a *diffeomorphic mapping* (or *diffeomorphism*) if it is smooth, bijective, and the inverse mapping φ^{-1}: $V \to U$ is also smooth. (When $r=0$, a bijective continuous mapping whose inverse mapping is also continuous is called a *homeomorphism*).

It follows directly from (2) that the *Jacobian $D\varphi$ of an arbitrary diffeomorphism φ: $U \to V$ is nonzero at all points of U* (with $(D\varphi^{-1})(\mathbf{y}) = (D\varphi)(\mathbf{x})^{-1}$ for any point $\mathbf{y} = \varphi(\mathbf{x})$ of V).

The **inverse mapping theorem** states that *if the Jacobian $D\varphi$ of a smooth mapping φ: $U \to V$ is nonzero at the point $x_0 \in U$, then there is an open set $U' \subset U$ which contains the point $\mathbf{x}_0$ such that the restriction $\varphi|_{U'}$ of φ to U' is a diffeomorphism of U' onto some open set $V' \subset V$ that contains a point $\mathbf{y}_0 = \varphi(\mathbf{x}_0)$.*

It follows, in particular, that a *smooth bijective mapping φ: $U \to V$ whose Jacobian is everywhere nonzero is a diffeomorphism.*

Remark 1. There are smooth nonbijective mappings φ: $U \to V$ with the Jacobian other than zero everywhere. The mapping φ of a plane ring $1 < x^2 + y^2 < 2$ onto itself, given by $\varphi(x, y) = (x^2 - y^2, 2xy)$ may serve as an example.

Now we are in a position to give basic definitions of "abstract cartography".

Let $\mathscr{X}$ be an arbitrary set.

Definition 1. A *chart* in $\mathscr{X}$ is a pair (U, h), where U is a subset in $\mathscr{X}$ and h is a mapping of the set U into $\mathbb{R}^n$ bijectively representing U on some open set of $\mathbb{R}^n$. The set U is called the *domain of definition* (or *support*) of the chart (U, h) and h is the *chart mapping*. For any point $p \in U$ a point $h(p) \in \mathbb{R}^n$ is of the form $(x^1(p), \ldots, x^n(p))$, where $x^1(p), \ldots x^n(p) \in \mathbb{R}$. This gives n numerical functions:

$$(3) \quad x^1\colon p \mapsto x^1(p), \ldots, x^n\colon p \mapsto x^n(p), \; p \in U,$$

on U which are called the *local coordinates* of a chart (U, h).

Since coordinates (3) uniquely define the mapping h, we usually write $(U, x^1, \ldots, x^n)$ rather than (U, h) and h is called a *coordinate mapping.*

Remark on notation.

By $\varphi(x^1, \ldots, x^n)$, where φ is some function on $h(U)$, we shall denote the function

$$\varphi \circ h\colon p \mapsto \varphi(x^1(p), \ldots, x^n(p)) \text{ on } U,$$

which is in full agreement with the traditional notation of a composite function. Thus

$$y = \varphi (x^1, \ldots, x^n)$$

implies that y is just another way of denoting the function $\varphi (x^1, \ldots, x^n)$. To reduce the number of letters used the function φ will sometimes be also denoted by y, and hence we shall write

$$y = y (x^1, \ldots, x^n).$$

Notice that in this formula the letter y designates **different functions** on the right and on the left. On the left, it is a function on U, and on the right it is a function on $h (U)$ which serves to express functions on U in terms of functions $x^1, \ldots, x^n$. When these functions should be distinguished, the first of them will be denoted by $y \circ h$.

We stress that we shall never, contrary to the common practice, use $\varphi (x^1, \ldots, x^n)$ (or $y (x^1, \ldots, x^n)$) to denote the function φ (the function y) on $h (U)$ (and only sometimes we will denote by the symbol the value of this function at the point $(x^1, \ldots, x^n) \in h (U)$).

Two charts (U, h) and (V, k) are said to be *overlapping* if $U \cap V \neq \varnothing$. (This terminology reflects the general tendency, we are going to follow sometimes, not to distinguish pedantically between (U, k) and U.)

Let (U, k) and (V, h) be two overlapping charts, and let $W = U \cap V$. Then defined in $\mathbb{R}^n$ are the two sets $h (W)$ and $k (W)$ and the mapping

$$(k \mid_W) \circ (h \mid_W)^{-1}: h (W) \to k (W) \tag{4}$$

of the first set onto the second. Speaking loosely, we shall denote mapping (4) by $k \circ h^{-1}$.

Definition 2. Two charts (U, h) and (V, k) in $\mathscr{X}$ are said to be *compatible* if they either do not overlap ($U \cap V = \varnothing$), or:

(a) both sets $h (W)$ and $k (W)$, where $W = U \cap V$, are open in $\mathbb{R}^n$,

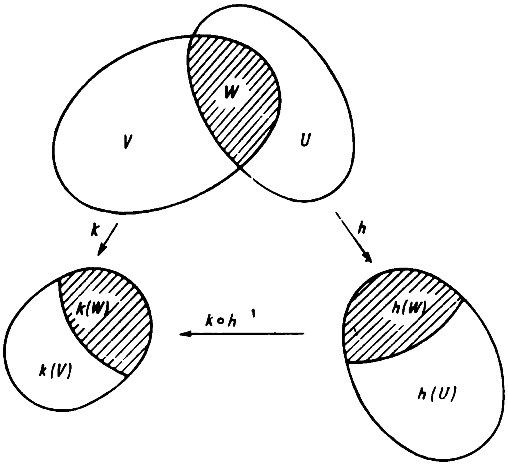
(b) mapping (4) is a diffeomorphism (a homeomorphism when $r = 0$).

If $\varphi^1, \ldots, \varphi^n$ are functions defining diffeomorphism (4), then for the restrictions $x^1 \mid_W, \ldots, x^n \mid_W$ and

$y^1 |_W, \ldots, y^n |_W$ of the local coordinates of charts (U, h) and (V, k) on W the formulas

$$(5) \qquad \begin{aligned} y^1|_W &= \varphi^1 (x^1|_W, \ldots, x^n|_W), \\ &\cdots\cdots\cdots\cdots\cdots \\ y^n|_W &= \varphi^n (x^1|_W, \ldots, x^n|_W) \end{aligned}$$

will hold. Due to this, $\varphi^1, \ldots, \varphi^n$ are called the *functions of transition* (on W) from the coordinates $x^1, \ldots, x^n$ to $y^1, \ldots, y^n$, and formulas (5) are called *transition formulas*. The diffeomorphism $k \circ h^{-1}$ is called the *transition mapping*. As a rule, the subscript W in formulas (5) is omitted and they are written as



Graphically the sets h(U) and k(V) do not intersect though generally this may not be the case. The sets h (U), h (W) and k (V) may be not connected either

$$(6) \quad y^1 = \varphi^1 (x^1, \ldots, x^n), \ldots, y^n = \varphi^n (x^1, \ldots x^n)$$

(sometimes adding "on W").

We shall write formulas (5) in a shorter ("vector") form

$$(7) \qquad \mathbf{y} = \varphi (\mathbf{x}) \ (\text{or } \mathbf{y} = (k \circ h^{-1})\, \mathbf{x}).$$

We should understand that formulas (7) are just a short way of writing formulas (6) or (5).

Although formulas (7) have the form of relations between the points of some subsets of $\mathbb{R}^n$, in fact they connect not the points of $\mathbb{R}^n$ but functions given on a subset of $\mathscr{X}$ (and in this sense they are relations in $\mathscr{X}$).

Definition 3. A set of charts $\{(U_\alpha, h_\alpha)\}$ is said to be an *atlas on* $\mathscr{X}$ if

(a) any two charts of that set are compatible,

(b) an equation

$$\bigcup_\alpha U_\alpha = \mathscr{X}$$

holds (charts (U_α, h_α) cover the whole of $\mathscr{X}$).

For any atlas A we denote by $\mathsf{A}_{\max}$ the set of all charts compatible with each chart of A.

Proposition 1. *The set* $\mathsf{A}_{\max}$ *is an atlas.*

Proof. Let (U, h) and (V, k) be two overlapping charts in $\mathsf{A}_{\max}$, and let $\mathbf{x}$ be an arbitrary point of a set $h(W)$, where as before $W = U \cap V$.

Consider a point $p = h^{-1}(\mathbf{x}) \in W$. Since the set $\mathsf{A} = \{(U_\alpha, h_\alpha)\}$ is an atlas, there is a chart (U_α, h_α) in A such that $p \in U_\alpha$. Since under the hypothesis (U, h) and (V, k) are compatible with (U_α, h_α), the sets $h_\alpha(U_\alpha \cap U)$ and $h_\alpha(U_\alpha \cap V)$ are open in $\mathbb{R}^n$. Therefore their intersection

$$h_\alpha(U_\alpha \cap U) \cap h_\alpha(U_\alpha \cap V) = h_\alpha(U_\alpha \cap W)$$

is open in $\mathbb{R}^n$ (and hence in $h_\alpha(U_\alpha \cap U)$). Moreover, the set $h(U_\alpha \cap U)$ is also open in $\mathbb{R}^n$ and the mapping

$$h \circ h_\alpha^{-1}\colon h_\alpha(U_\alpha \cap U) \to h(U_\alpha \cap U)$$

is a diffeomorphism (and therefore, being in particular a homeomorphism, carries open sets into open sets). Therefore the set

$$(h \circ h_\alpha^{-1})(h_\alpha(U_\alpha \cap W)) = h(U_\alpha \cap W)$$

is open (in $h(U_\alpha \cap U)$ and hence in $\mathbb{R}^n$). Since $\mathbf{x} \in h(U_\alpha \cap W) \subset h(W)$, this proves that $\mathbf{x}$ is an interior point of the set $h(W)$, and since x was an arbitrary point in $h(W)$ this proves that the set $h(W)$ is open.

Interchanging the roles played by h and k, we shall similarly prove that $k(W)$ is also open.

Further, since

$$k \circ h^{-1} = (k \circ h_\alpha^{-1}) \circ (h_\alpha \circ h^{-1}) \tag{8}$$

and both mappings $k \circ h_\alpha^{-1}$ and $h_\alpha \circ h$ are diffeomorphisms, the mapping $k \circ h^{-1}$ is also a diffeomorphism (as a composition of diffeomorphisms). Hence (U, h) and (V, k) are compatible charts.

In fact the conclusion that $k \circ h^{-1}$ is a diffeomorphism is somewhat hasty, since strictly speaking formula (8) is meaningless. Indeed, the mapping $h_\alpha \circ h^{-1}$ (and more precisely the mapping $(h_\alpha |_{U_\alpha \cap U}) \circ (h \mid_{U_\alpha \cap U})^{-1})$ is a mapping from $h(U_\alpha \cap U)$ onto $h_\alpha(U_\alpha \cap U)$ and $k \circ h_\alpha^{-1}$ (more precisely, $(k \mid_{U_\alpha \cap V}) \circ (h_\alpha \mid_{U_\alpha \cap V})^{-1})$ is a mapping from $h_\alpha(U_\alpha \cap V)$ onto $k(U_\alpha \cap V)$, and so we have no right to consider the composition of these mappings. To get the right, we must restrict all mappings h, k, and h_α to $U_\alpha \cap W$, i.e. write formula (8) in the following, now unimpeachable, form:

$$\begin{aligned}&(k|_{U_\alpha \cap W}) \circ (h_{U_\alpha \cap W})^{-1}\\ &\quad= [(k|_{U_\alpha \cap W}) \circ (h_\alpha|_{U_\alpha \cap W})^{-1}] \circ [(h_\alpha|_{U_\alpha \cap W}) \circ (h_{U_\alpha \cap W})^{-1}].\end{aligned}$$

But then we can only state that a diffeomorphism is the mapping

$$(k|_{U_\alpha \cap W}) \circ (h|_{U_\alpha \cap W})^{-1} = [(k|_W) \circ (h|_W)^{-1}]|_{h(U_\alpha \cap W)}, \tag{9}$$

and to go over to the entire mapping

$$k \circ h^{-1} = (k \mid_W) \circ (h \mid_W)^{-1} \tag{10}$$

additional considerations are necessary. These considerations are trivial enough, however. Namely, since mapping (9) is a diffeomorphism, its Jacobian at x is nonzero (we assume that $r > 0$). But it is clear that at x the Jacobians of mappings (9) and (10) take on the same values (for in the neighbourhood $h(U_\alpha \cap W)$ of x these mappings coincide). Thus the Jacobian of mapping (10) at an arbitrary point $\mathbf{x}$ of an open set $h(W)$ is nonzero. Hence, since the mapping is smooth and bijective, this mapping is a diffeomorphism. (When $r = 0$, instead of referring to a Jacobian, it is necessary to make use of the fact that a mapping is continuous if and only if it is continuous at every point.) In what follows this sort of refinements will be left to the reader.

So we have proved that any two charts in $\mathsf{A}_{\max}$ are compatible, i.e. that this set satisfies condition (a) of Definition 3. Since condition (b) of Definition 3 obviously holds for $\mathsf{A}_{\max}$ (it holds even for $\mathsf{A} \subset \mathsf{A}_{\max}$), this completes the proof of Proposition 1. □

Clearly if A and A^* are atlases and $\mathsf{A} \subset \mathsf{A}^*$, then $\mathsf{A}^*_{\max} \subset \mathsf{A}_{\max}$. Therefore, in particular $(\mathsf{A}_{\max})_{\max} \subset \mathsf{A}_{\max}$, and hence $(\mathsf{A}_{\max})_{\max} = \mathsf{A}_{\max}$. Thus if $\mathsf{A}_{\max} \subset \mathsf{A}^*$, then $\mathsf{A}^*_{\max} \subset (\mathsf{A}_{\max})_{\max} = \mathsf{A}_{\max}$ and therefore $\mathsf{A}^* \subset \mathsf{A}_{\max}$. This means that the *atlas* $\mathsf{A}_{\max}$ *is maximal* (in the atlase set partially ordered with respect to inclusion). In addition, if $\overline{\mathsf{A}}$ is an arbitrary maximal atlas containing the atlas A, then $\overline{\mathsf{A}} \subset \overline{\mathsf{A}}_{\max} \subset \mathsf{A}_{\max}$, and hence $\overline{\mathsf{A}} = \mathsf{A}_{\max}$. This proves

Corollary 1. *Every atlas* A *is contained in a unique maximal atlas* $\mathsf{A}_{\max}$. □

Now we are in a position to give our main definition.

Definition 4. Maximal atlases on $\mathscr{X}$ are also called *smooth structures* (or simply *smoothnesses*). A set $\mathscr{X}$ with a smooth structure $\mathsf{A}_{\max}$ given on it is called a *smooth manifold*. (Thus smooth manifolds are, strictly speaking, pairs of the form $(\mathscr{X}, \mathsf{A}_{\max})$, but to simplify this notation we shall use it in full when this cannot be done without.) Charts of an atlas $\mathsf{A}_{\max}$ are called *charts of the manifold* $\mathscr{X}$ or even, for greater expressiveness, *smooth charts* of the atlas.

By definition two manifolds $(\mathscr{X}, \mathsf{A}_{\max})$ and $(\mathscr{Y}, \mathsf{A}^*_{\max})$ are the same if and only if $\mathscr{X} = \mathscr{Y}$ and $\mathsf{A}_{\max} = \mathsf{A}^*_{\max}$.

Two atlases are said to be *equivalent* if they are contained in the same maximal atlas. Atlases A and A^* are equivalent if and only if their union $\mathsf{A} \cup \mathsf{A}^*$ is an atlas (i.e. each chart of any of these atlases is compatible with each chart of the other).

To specify a smoothness $\mathsf{A}_{\max}$ on $\mathscr{X}$ it suffices to give an arbitrary atlas $\mathsf{A} \subset \mathsf{A}_{\max}$. Thus pairs of the form $(\mathscr{X}, \mathsf{A})$, where A is an arbitrary atlas on $\mathscr{X}$, may be assumed to be smooth manifolds. Manifolds $(\mathscr{X}, \mathsf{A})$ and $(\mathscr{Y}, \mathsf{A}^*)$ will be the same if and only if $\mathscr{X} = \mathscr{Y}$ and A and A^* are equivalent (i.e. if their union $\mathsf{A} \cup \mathsf{A}^*$ is an atlas).

It should be stressed that our definition of a smooth manifold contains the number n, the dimension of the space $\mathbb{R}^n$ which contains the images $h(U)$ of the supports of charts.

Definition 5. This number is called the *dimension* of a smooth manifold $\mathscr{X}$ and is denoted by $\dim \mathscr{X}$.

The definition of a smooth manifold $\mathscr{X}$ also contains a number $r \geqslant 0$ (or symbols ∞ or ω), the smoothness class of transition mappings $k \circ h^{-1}$. This number is called the *smoothness class of a manifold* $\mathscr{X}$ (it is also said that $\mathscr{X}$ is a *C^r-manifold*). C^ω-manifolds are also known as *real-analytic manifolds.*

When $r = 0$ the term "smooth manifold" is not used and is replaced by the term *topological manifold.*

Of course, every C^r-manifold may be regarded directly as a $C^{r'}$-manifold for any $r' < r$ (and in particular as a topological manifold).

In accordance with the foregoing all manifolds being considered will be assumed to belong to the C^∞-class.

Examples of smooth manifolds.

Example 1. On a space $\mathbb{R}^n$ a pair $(\mathbb{R}^n, \mathrm{id})$, where $\mathrm{id}\colon \mathbb{R}^n \to \mathbb{R}^n$ is an identity mapping $\mathbf{x} \mapsto \mathbf{x}$, is a chart, and a one-element set of charts consisting of this chart is an atlas. The corresponding smooth structure on $\mathbb{R}^n$ (the C^ω-class) is called *standard.* In what follows we shall always consider $\mathbb{R}^n$ to be a smooth manifold with a standard smooth structure. Notice that in this structure $\dim \mathbb{R}^n = n$.

Problem 1. Prove that charts of a standard smooth structure on $\mathbb{R}^n$ are pairs (U, h), where U is an arbitrary open set in $\mathbb{R}^n$ and h is an arbitrary diffeomorphism of U onto some open set $h(U) \subset \mathbb{R}^n$.

Example 2. More generally, for any open set O of $\mathbb{R}^n$ a pair (O, id) is a chart and a one-element set of charts that consists of that chart is an atlas. The smooth structure on O defined by that atlas is said to be *induced* by the standard smooth structure on $\mathbb{R}^n$. Its charts are clearly the charts (U, h) of the latter structure for which $U \subset O$. In what follows all open sets $O \subset \mathbb{R}^n$ will be assumed to be smooth manifolds with that smooth structure.

Example 3. There are also nonstandard smooth structures on a space $\mathbb{R}$. Consider, for example, on $\mathbb{R}^1 = \mathbb{R}$ a chart $(\mathbb{R}, h_0)$, where h_0 is a mapping $\mathbb{R} \to \mathbb{R}$ given by $h_0(t) = t^3$, $t \in \mathbb{R}$. This chart also forms a one-element atlas, but (for $r > 0$) the corresponding smooth structure is nonstandard since the charts $(\mathbb{R}, \text{id})$ and $(\mathbb{R}, h_0)$ are not compatible (the transition mapping $h_0 \circ (\text{id})^{-1} = h_0$ is smooth and bijective, but the inverse mapping h_0^{-1}: $t \mapsto t^{1/3}$ is not differentiable when $t = 0$).

Example 4. For every parametrization $\gamma: I \to \mathscr{A}$, $I = (a, b)$, of an open simple regular arc $\mathscr{L}$ the pair $(\mathscr{L}, \gamma_0^{-1})$, where $\gamma^{-1}: \mathscr{L} \to I$ is the inverse of the mapping γ (considered as a mapping $I \to \mathscr{L}$) is a chart on $\mathscr{L}$, any two such charts being compatible according to Proposition 1 of Lecture 1. This means that *any open regular simple arc is naturally a one-dimensional manifold.*

Example 5. Similarly, any elementary surface $\mathscr{X}$ is a two-dimensional smooth manifold with charts of the form $(\mathscr{X}, \gamma^{-1})$, where γ is an arbitrary parametrization of $\mathscr{X}$.

Example 6. The simplest manifold which cannot be covered by one chart is a circle

$$\mathbb{S}^1: x^2 + y^2 = 1.$$

Let p_0 and q_0 be its points $(-1, 0)$ and $(1, 0)$, and let $U = \mathbb{S}^1 \setminus \{p_0\}$ and $V = \mathbb{S}^1 \setminus \{q_0\}$. For any point $p \in U$ (any point $q \in V$), we denote by $h(p)$ (by $k(p)$) the angle on $(-\pi, \pi)$ formed by the radius vector of that point and the positive (negative) direction of the abscissa axis. It is clear that the resulting mappings

$$h\colon U \to (-\pi, \pi), \quad k\colon V \to (-\pi, \pi)$$

are bijective, i.e. pairs (U, h) and $(V\ k)$ are charts on $\mathbb{S}^1$ (with $n = 1$). We have $U \cap V = \mathbb{S}^1 \setminus \{p_0, q_0\}$, the sets $h(U \cap V) = (-\pi, 0) \cup (0, \pi)$ and $k(U \cap V) = (-\pi, 0) \cup (0, \pi)$ (turning out to be equal) are open in $\mathbb{R}$, and the mapping

$$k \circ h^{-1}\colon h(U \cap V) \to k(U \cap V)$$

is given by

$$(k \circ h^{-1})(t) = \begin{cases} t + \pi & \text{if } t \in (-\pi, 0), \\ 1 - \pi & \text{if } t \in (0, \pi), \end{cases}$$

and is therefore a diffeomorphism.

Consequently, charts (U, h) and (V, k) are compatible, and since $U \cup V = \mathbb{S}^1$, they make up an atlas and hence define some (C^ω-) smoothness on $\mathbb{S}^1$.

Example 7. Similarly, it can be proved that an arbitrary closed (with no end points) simple regular arc in an affine space $\mathscr{A}$ is a one-dimensional manifold that can be covered by two charts.

Example 8. Let $U^{(+)}$, $U^{(-)}$, $V^{(+)}$, $V^{(-)}$ be the subsets of a circle $\mathbb{S}^1$ (*open semicircles*) consisting of points $p = (x, y)$, for which $y > 0$, $y < 0$, $x > 0$, $x < 0$ respectively. Also let $h^{(+)}$, $h^{(-)}$, $k^{(+)}$, $k^{(-)}$ be mappings of these sets in $\mathbb{R}$ that act by formulas $(x, y) \mapsto x$, $(x, y) \mapsto x$, $(x, y) \mapsto y$, $(x, y) \mapsto y$ (which are restrictions of projections onto coordinate axes). Each of the mappings represents bijectively the corresponding set on an open interval $(-1, 1)$ of the axis $\mathbb{R}$, and hence the pairs

$$(U^{(+)}, h^{(+)}), \quad (U^{(-)}, h^{(-)}), \quad (V^{(+)}, k^{(+)}), \quad (V^{(-)}, k^{(-)})$$

are the charts in $\mathbb{S}^1$ (x being a local coordinate on charts $U^{(\pm)}$ and y on charts $V^{(\pm)}$). These charts are easily seen to be compatible (any two of them either do not overlap or their intersection is a quarter of a circle projected onto open intervals $(0, 1)$ or $(-1, 0)$ of the axis $\mathbb{R}$, the corresponding transition mappings being given by formulas of the form $y = \pm \sqrt{1 - x^2}$ and hence regarded as diffeomorphisms) and cover the entire circle $\mathbb{S}^1$, i.e. make up an atlas on S^1. *This atlas is equivalent to the atlas of Example* 6 (and therefore defines the same smoothness on S^1). Indeed, say, for charts (U, h) and $(U^{(+)}, h^{(+)})$ the intersection $U \cap U^{(+)}$ coincides with $U^{(+)}$, with $h(U^{(+)}) = (0, \pi)$, $h^{(+)}(U^{(+)}) = (-1, 1)$ and the mapping

$$h^{(+)} \circ h^{-1}\colon\ (0, \pi) \to (-1, 1)$$

being given by $x = \cos t$, $0 < t < \pi$ and hence acting as a diffeomorphism (for the remaining pairs of charts the situation is similar).

Example 9 (extension of Example 8 to an arbitrary dimension). Similarly, on a unit n-dimensional $(n \geqslant 1)$ sphere

$$S^n\colon\ x_0^2 + x_1^2 + \ldots + x_n^2 = 1$$

of a space $\mathbb{R}^{n+1}$ the pairs $(U_i^{(+)}, h_i^{(+)})$ and $(U_i^{(-)}, h_i^{(-)})$, $i = 0, 1, \ldots, n$, where $U_i^{(+)}$ and $U_i^{(-)}$ are *hemispheres* defined by $x_i > 0$ and $x_i < 0$, respectively, and $h_i^{(+)}$ and $h_i^{(-)}$ are restrictions to these hemispheres projected onto the i-th coordinate plane $x_i = 0$, are charts, and any two of these charts are compatible. For example, the mapping $h_i^{(+)}$, which is the mapping of a hemisphere $U_i^{(+)}$ onto a ball $\mathring{\mathbf{B}}^n\colon t_1^2 + \ldots + t_n^2 < 1$ of $\mathbb{R}^n$ (with coordinates $t_1, \ldots, t_n$), is given as follows:

$$t_a = \begin{cases} x_{a-1} & \text{if } a \leqslant i, \\ x_a & \text{if } a > i, \end{cases} \qquad a = 1, \ldots, n$$

(so that the numbers $t_1, \ldots, t_n$, i.e. the numbers $x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$, are local coordinates of the chart $(U_i^{(+)}, h_i^{(+)})$), and the inverse mapping $(h_i^{(+)})^{-1}\colon \mathring{\mathbf{B}}^n \to U_i^{(+)}$ is given as follows:

$$x_b = \begin{cases} t_{b+1} & \text{if } b < i, \\ \sqrt{1 - \mathbf{t}^2} & \text{if } b = i, \\ t_b & \text{if } b > i, \end{cases} \qquad b = 0, 1, \ldots, n,$$

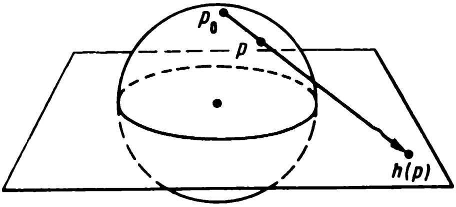
where $\mathbf{t}^2 = t_1^2 + \ldots + t_n^2$. For $i \leqslant j$ therefore the mapping $h_j^{(+)} \circ (h_{i-1}^{(+)})^{-1}$ will be defined on a hemisphere $\mathring{\mathbf{B}}^n_{j>0} = \{\mathbf{t} \in \mathring{\mathbf{B}}^n;\ \ t_j > 0\}$ and, mapping the hemisphere onto the half-ball $\mathring{\mathbf{B}}^n_{i>0} = \{\mathbf{t} \in \mathring{\mathbf{B}}^n;\ t_i > 0\}$, will be given as follows:

$$(11)\qquad t'_a = \begin{cases} t_a & \text{if } a < i, \\ \sqrt{1-\mathbf{t}^2} & \text{if } a = 1, \\ t_{a-1} & \text{if } i \leqslant a \leqslant j, \\ t_a & \text{if } a > j. \end{cases} \qquad a = 1, \ldots, n.$$

Since these formulas obviously specify a diffeomorphism, the charts $(U^{(+)}_{i-1}, h^{(+)}_{i-1})$ and $(U^{(+)}_j, h^{(+)}_j)$ are consequently compatible. For the other pairs of charts in (11) subscripts may be shifted and a minus sign may appear before the radical, which exerts no influence, of course, on the final conclusion that we have a diffeomorphism.

Since the charts constructed obviously cover the entire sphere $\mathbb{S}^n$, they make up an atlas and hence define some smoothness on $\mathbb{S}^n$.

Example 10. The atlas constructed in Example 9 consists of $2n + 2$ charts. It turns out that the same smoothness can be given on $\mathbb{S}^n$ by an atlas consisting of but two charts (cf. Example 6). To this end, for any point $p = (x_0, x_1, \ldots, x_n)$ of $\mathbb{S}^n$ other than the point $p_0 = (1, 0, \ldots, 0)$ (the *north pole* of $\mathbb{S}^n$) we consider a straight line p_0p in the space $\mathbb{R}^{n+1}$. The canonical equations of that straight line are of the form



A stereographic projection

$$\frac{X_0 - 1}{x_0 - 1} = \frac{X_1}{x_1} = \ldots = \frac{X_n}{x_n},$$

and hence the straight line intersects the hyperplane $X_0 = 0$ (the *equatorial hyperplane* of $\mathbb{S}^n$) at a point with coordinates

$$t_1=\frac{x_1}{1-x_0}, \ \ldots, \ t_n=\frac{x_n}{1-x_0} \tag{12}$$

(the coordinates on the equatorial hyperplane will be denoted by $t_1, \ldots, t_n$). Point (12) is called a *stereographic projection* of a point $p \in \mathbb{S}^n \setminus \{p_0\}$ (cf. Lecture I.28). Denoting this point by $h(p)$ we obtain a mapping h (obviously bijective) of a set $U = \mathbb{S}^n \setminus \{p_0\}$ onto a space $\mathbb{R}^n$, i.e. some chart (U, h) (with $h(U) = \mathbb{R}^n$).

Similarly, proceeding from the *south pole* $q_0 = (-1, 0, \ldots, 0)$ of $\mathbb{S}^n$, we can construct a stereographic projection $k: V \to \mathbb{R}^n$, where $V = \mathbb{S}^n \setminus \{q_0\}$, carrying an arbitrary point $p = (x_0, x_1, \ldots, x_n)$ of $\mathbb{S}^n$ other than the point q_0 into a point of the equatorial plane with coordinates

$$t'_1=\frac{x_1}{1+x_0}, \ \ldots, \ t'_n=\frac{x_n}{1+x_0}, \tag{13}$$

and hence obtain a chart (V, k).

For the charts constructed the intersection $W = U \cap V$ is of the form $\mathbb{S}^n \setminus \{p_0, q_0\}$, and each mapping h or k represents this intersection on an open set $\mathbb{R}^n_* = \mathbb{R}^n \setminus \{0\}$ of a space $\mathbb{R}^n$. The corresponding transition mapping

$$k \circ h^{-1}: \mathbb{R}^n_* \to \mathbb{R}^n_* \tag{14}$$

will be given by the formulas obtained by eliminating $x_0, x_1, \ldots, x_n$ from (12) and (13). But according to formulas (12) $x_i = (1 - x_0)\, t_i$, and hence

$$\begin{aligned} 1 &= x_0^2 + x_1^2 + \ldots + x_n^2 \\ &= x_0^2 + (1 - x_0)^2\, (t_1^2 + \ldots + t_n^2) \\ &= x_0^2 + (1 - x_0)^2\, \mathbf{t}^2, \end{aligned}$$

i.e.

$$(1 + \mathbf{t}^2)\, x_0^2 - 2\mathbf{t}^2 x_0 + \mathbf{t}^2 - 1 = 0.$$

This equation has a root $x_0 = 1$, which is of no use for us (it corresponds to the eliminated point p_0), and for the second root of the equation

$$x_0=\frac{\mathbf{t}^2-1}{1+\mathbf{t}^2}$$

we have

$$1-x_0=\frac{2}{1+\mathbf{t}^2}, \quad 1+x_0=\frac{2\mathbf{t}^2}{1+\mathbf{t}^2}.$$

For any $i = 1, \ldots, n$ therefore

$$x_i=(1-x_0)\,t_i=\frac{2t_i}{1+\mathbf{t}^2}$$

and hence

$$t'_1=\frac{x_i}{1+x_0}=\frac{t_i}{\mathbf{t}^2}.$$

This proves that mapping (14) can be written as

$$t'_1=\frac{t_1}{\mathbf{t}^2}, \ \ldots, \ t'_n=\frac{t_n}{\mathbf{t}^2}$$

and is therefore a diffeomorphism.

Consequently, charts (U, h) and (V, k) are compatible and so make up an atlas.

Problem 2. Prove that this atlas is equivalent to the atlas of Example 9 and hence defines on the same smoothness a sphere S^n.

Example 11. Recall that by an n-dimensional space $\mathbb{R}P^n$ is meant the set of all classes $(x_0 : x_1 : \ldots : x_n)$ of proportional vectors $(x_0, x_1, \ldots, x_n) \neq 0$ or which is the same the set of all straight lines of the space $\mathbb{R}^{n+1}$ through the point 0. Let U_i, $i = 0, 1, \ldots, n$, be the set of all points $p = (x_0 : x_1 : \ldots : x_n)$ of a space $\mathbb{R}P^n$ for which $x_i \neq 0$, and let h_i be a mapping $U_i \to \mathbb{R}^n$ carrying a point $p = (x_0 : x_1 : \ldots : x_n)$ of U_i into a point $(t_1, \ldots, t_n)$ of $\mathbb{R}^n$ for which

$$t_a=\begin{cases}\dfrac{x_{a-1}}{x_i} & \text{if } a\leqslant i,\\[2mm] \dfrac{x_a}{x_i} & \text{if } a>i,\end{cases} \qquad a=1, \ldots, n. \tag{15}$$

It is clear that the mapping h_i is bijective, so that the pair (U_i, h_i), $i = 0, 1, \ldots, n$, is a chart in $\mathbb{R}P^n$.

For any subscripts i and $j \neq i$ the intersection $(U_i \cap U_j)$ consists of points $p = (x_0 : x_1 : \ldots : x_n)$ for which $x_i \neq 0$ and $x_j \neq 0$, and the sets $h_i(U_i \cap U_j)$ and $h_j(U_i \cap U_j)$ consist (for $j > i$) of points $\mathbf{t} = (t_1, \ldots, t_n)$

of $\mathbb{R}^n$ for which $t_j \neq 0$ and $t_{i+1} \neq 0$ respectively. It is clear that these sets are open.

Since a mapping h_i acts by formulas (15), and h_j by similar formulas obtained by replacing i and j, the transition mapping $h_j \circ h_i^{-1}$ carries an arbitrary point $(t_1, \ldots, t_n) \in h_i(U_i \cap U_j)$ into a point $(t'_1, \ldots, t'_n) \in h_j(U_i \cap U_j)$ for which

$$t'_a = \begin{cases} \dfrac{t_a}{t_j} & \text{if } a \leqslant i \text{ or } a > j, \\ \dfrac{1}{t_j} & \text{if } a = i+1, \\ \dfrac{t_{a-1}}{t_j} & \text{if } i+1 < a \leqslant j \end{cases} \qquad a = 1, \ldots, n,$$

and is thus a diffeomorphism.

This proves that any two of the charts (U_i, h_j) and (U_j, h_j) are compatible, and hence these charts define a smooth structure on $\mathbb{R}P^n$.

Notice that in all the Examples (except, naturally, for Examples 4, 5, 7) real analytic (C^ω-) manifolds resulted.

Lecture 7

Topology of a smooth manifold· Open submanifolds· Neighbourhoods and interior points· Homeomorphisms· The first axiom of countability and the property of being locally flat· The second axiom of countability· Non-Hausdorff manifolds· Smoothnesses of a topological space· Topological manifolds· Zero-dimensional manifolds· The category TOP· The category DIFF· Pullback of smoothness

---

Let $\mathscr{X}$ be an arbitrary smooth (or topological) manifold.

Definition 1. A subset $O \subset \mathscr{X}$ is said to be *open* (in $\mathscr{X}$) if for any chart (U, h) of $\mathscr{X}$ the set $h(O \cap U) \subset \mathbb{R}^n$ is open (in $\mathbb{R}^n$).

It is easy to see that for $\mathscr{X} = \mathbb{R}^n$ this definition gives ordinary open sets in $\mathbb{R}^n$.

Open sets in the space $\mathscr{X} = \mathbb{R}^n$ (as well as in any metric space $\mathscr{X}$) are known from calculus to have the following properties:

1° an open set $\varnothing$ and the entire space $\mathscr{X}$ are open,

2° the union of an arbitrary family of open sets is open,

3° the intersection of an arbitrary finite family of open sets is open.

Since

$$\bigcup_\alpha h(O_\alpha \cap U) = h\left(\left(\bigcup_\alpha O_\alpha\right) \cap U\right)$$

and

$$\bigcap_\alpha h(O_\alpha \cap U) = h\left(\left(\bigcap_\alpha O_\alpha\right) \cap U\right)$$

for any chart (U, h) and any family $\{O_\alpha\}$ of subsets of $\mathscr{X}$, open sets have the same properties in $\mathscr{X}$.

Definition 2. A set $\mathbb{T}$ of subsets of the set $\mathscr{X}$ is said to be a *topology* (or a *topological structure*) on $\mathscr{X}$ and the sets of $\mathbb{T}$ *open sets* if these sets have Properties 1° to 3°. The set $\mathscr{X}$, together with topology $\mathbb{T}$ (i.e., strictly speaking, a pair $(\mathscr{X}, \mathbb{T})$) specified on it, is called a *topological space*.

We can thus say that Definition 1 introduces some topology into every manifold or, to put it otherwise, that by virtue of this definition *every manifold is a topological space*. In this regard the topology on $\mathscr{X}$ is a structure derived from the smoothness on $\mathscr{X}$.

This situation is quite similar to that for metric spaces whose topology is a structure derived from a metric.

In what follows, when speaking of a topology on a smooth (or topological) manifold we shall always have in mind the topology introduced by Definition 1.

According to Definition 1, in order to see whether or not a given subset $O \subset \mathscr{X}$ is open, it is necessary to consider the sets $h(O \cap U)$ for all charts (U, h) of $\mathscr{X}$. In practice this is impossible, of course, and the question arises as to whether it is possible to restrict somehow the number of charts. It turns out that it is quite enough to take the charts of one arbitrary atlas, and what is more, only the charts covering the set O. That is, the following proposition is true.

Proposition 1. *Let* $\{(U_\alpha, h_\alpha)\}$ *be a family of charts of the manifold* $\mathscr{X}$ *such that*

$$O \subset \bigcup_\alpha U_\alpha.$$

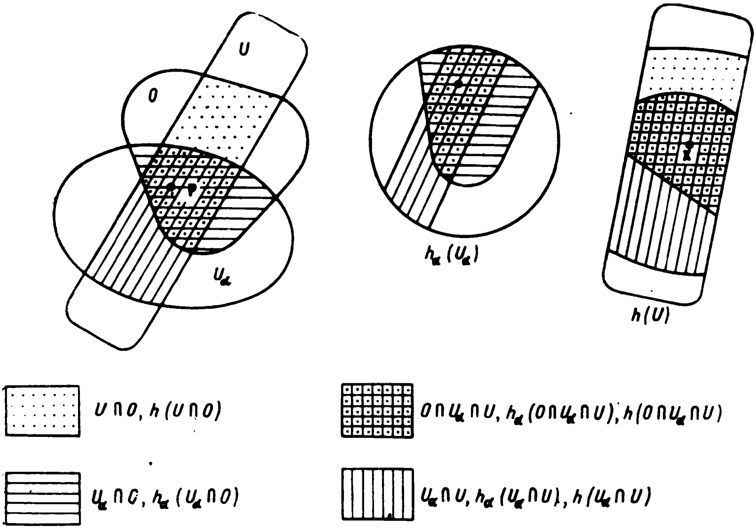
Then, for O *to be an open set in* $\mathscr{X}$, *it is sufficient that for any* α *the set* $h_\alpha(O \cap U_\alpha)$ *is open in* $\mathbb{R}^n$.

Proof. It is necessary to prove that if the set O satisfies the conditions of the proposition, then for any chart (U, h) of the manifold $\mathscr{X}$ the set $h(O \cap U)$ is open in $\mathbb{R}^n$, i.e., if the set is nonempty, any point $\mathbf{x}$ of the set is an interior point. With this in mind, consider a point $p = h^{-1}(\mathbf{x})$ in $\mathscr{X}$.

Since $\mathbf{x} \in h(O \cap U)$, there is $p \in O \cap U$ and, in particular, $p \in O$. There is α therefore such that $p \in U_\alpha$.

The charts (U_α, h_α) and (U, h) being compatible, the sets $h_\alpha(U_\alpha \cap U)$ and $h(U_\alpha \cap U)$ are open in $\mathbb{R}^n$. Moreover, under the hypothesis the set $h_\alpha(O \cap U_\alpha)$ is also open. Hence so is the set

$$h_\alpha(O \cap U_\alpha \cap U) = h_\alpha(O \cap U_\alpha) \cap h_\alpha(U_\alpha \cap U).$$


U
O
U_α
$h_\alpha(U_\alpha)$
$h(U)$
$U \cap O$, $h(U \cap O)$
$O \cap U_\alpha \cap U$, $h_\alpha(O \cap U_\alpha \cap U)$, $h(O \cap U_\alpha \cap U)$
$U_\alpha \cap O$, $h_\alpha(U_\alpha \cap O)$
$U_\alpha \cap U$, $h_\alpha(U_\alpha \cap U)$, $h(U_\alpha \cap U)$

On the other hand, the homeomorphism

$$h \circ h_\alpha^{-1}\colon h_\alpha(U_\alpha \cap U) \to h(U_\alpha \cap U)$$

carries this set into a set $h(O \cap U_\alpha \cap U)$ which is, consequently, also open. Since

$$\mathbf{x} \in h(O \cap U_\alpha \cap U) \subset h(O \cap U),$$

this proves that the point $\mathbf{x}$ is an interior point of the set $h(O \cap U)$. □

Corollary 1. *For any chart (U, h) of the manifold $\mathscr{X}$ the subset $V \subset U$ is open in $\mathscr{X}$ if and only if the set $h(V)$ is open in $\mathbb{R}^n$.*

In particular, the set U is itself open in $\mathscr{X}$. □

For any open subset O of an arbitrary (C^r-) manifold $\mathscr{X}$ all charts (U, h) for which $U \subset O$ obviously make up an atlas on O (see Example 2 of the preceding lecture). The atlas is maximal, i.e. it is a smoothness on O (of the same class C^r). This smoothness is said to be *induced* by the smoothness of the manifold $\mathscr{X}$ and the set O having this smoothness is called an *open submanifold* of $\mathscr{X}$. By definition $\dim O = \dim \mathscr{X}$.

In what follows every open set O of an arbitrary manifold $\mathscr{X}$ will always be considered to be a manifold with induced smoothness.

Notice that every atlas $\mathbf{A} = \{(U_\alpha, h_\alpha)\}$ of $\mathscr{X}$ defines an atlas $O \cap \mathbf{A}$ of O consisting of charts $(O \cap U_\alpha, h_\alpha |_{O \cap U_\alpha})$. Thus, to obtain an induced smoothness on O there is no need to consider all the charts on $\mathscr{X}$, it suffices to consider just the charts of one atlas.

It follows directly from Property 2° of open sets that for any subset A of a topological space $\mathscr{X}$ there is a maximal open set (possibly empty) contained in A (it is the union of all open sets contained in A). That open set is called the *interior* of A and is denoted by $\operatorname{Int} A$ (or $\overset{\circ}{A}$). Its points are called *interior points* of A.

Clearly, the *set A is open if and only if* $\operatorname{Int} A = A$, *i.e. if all its points are interior.*

For any point (or, more generally, for any subset) of a topological space $\mathscr{X}$ every open set containing this point (this subset) is called a *neighbourhood* of this point (this subset). By definition a *point is an interior point of a subset A if and only if some neighbourhood of that point is contained in A.*

Thus for $\mathscr{X} = \mathbb{R}^n$ this concept of an interior point coincides with that known from calculus.

For any chart (U, h) of an arbitrary manifold $\mathscr{X}$ the set U (the support of the chart) is a neighbourhood of every point $p \in U$. For this reason the supports of the charts of $\mathscr{X}$ are also called *coordinate neighbourhoods.*

It is clear that any neighbourhood V of a point $p \in U$ contained in a coordinate neighbourhood U is also a coordinate neighbourhood (with coordinate mapping $h|_V$) Hence *any neighbourhood O of p contains some coordinate neighbourhood* (such a neighbourhood is, for example, the intersection $O \cap U$).

The family of the neighbourhoods of a point p of a topological space $\mathscr{X}$ is called the *base for the neighbourhood system of the point* if any neighbourhood of p contains a neighbourhood from this family. (Intuitively speaking, a base must contain arbitrarily small neighbourhoods.)

In these terms the statement we have proved means that the *coordinate neighbourhoods of every point of an arbitrary manifold form a base for its neighbourhoods.*

A mapping $f: \mathscr{X} \to \mathscr{Y}$ of topological spaces is said to be a *homeomorphism* if it is bijective, and a set $O \subset \mathscr{X}$ is open in $\mathscr{X}$ if and only if the set $fO \subset \mathscr{Y}$ is open in $\mathscr{Y}$. Thus a homeomorphism is a bijective mapping which establishes a one-to-one correspondence between topologies.

Spaces $\mathscr{X}$ and $\mathscr{Y}$ are said to be *homeomorphic* if there is at least one homeomorphism $\mathscr{X} \to \mathscr{Y}$. Clearly, topological properties of homeomorphic spaces are the same (i.e. properties stated exclusively in terms of open sets). They are thus topologically indistinguishable.

An example of a topological property is that a topological space $\mathscr{X}$ has a countable base for the neighbourhood system of its points p. Such spaces are said to satisfy the *first axiom of countability* (or to be the first *countable (topological) spaces*).

It is easy to see that *any metric space $\mathscr{X}$ satisfies the first axiom of countability* (for any point of the space open balls with centre at that point whose radii are rational numbers form a countable base for the neighbourhood system of that point). In particular, the *space $\mathbb{R}^n$ satisfies the first axiom of countability.*

Every subset A of a topological space $\mathscr{X}$ has a topology (called an *induced topology*), whose open sets are the intersections $O \cap A$ of open sets O of $\mathscr{X}$ with A. A subset A provided with that topology is called a *subspace* of $\mathscr{X}$.

Clearly, any subspace of a space that satisfies the first axiom of countability also satisfies that axiom. Hence, in particular, *any subspace of space $\mathbb{R}^n$ satisfies the first axiom of countability.*

A family $\{U_\alpha\}$ of open sets of a topological space $\mathscr{X}$ is said to be its *open covering* if any point of that space is at least in one element of the family, i.e. if

$$\bigcup_\alpha U_\alpha = \mathscr{X}.$$

For example, the supports U_α of charts (U_α, h_α) of an arbitrary atlas of the manifold $\mathscr{X}$ make up an open covering of the manifold.

A space $\mathscr{X}$ is called *locally Euclidean* if it has an open covering $\{U_\alpha\}$ each element U_α of which is homeomorphic to some open set of $\mathbb{R}^n$ (and hence satisfies the first axiom of countability).

But clearly if a topological space $\mathscr{X}$ has an open covering $\{U_\alpha\}$ all of whose elements U_α satisfy (in induced topology) the first axiom of countability, then the space $\mathscr{X}$ itself also satisfies the axiom. (This fact is expressed by saying that the property of satisfying the first axiom of countability is a *local property*.) Hence *every locally Euclidean space satisfies the first axiom of countability.*

On the other hand, it follows directly from Corollary 1 that *for any chart* (U, h) *of an arbitrary manifold* $\mathscr{X}$ *the mapping* $h: U \rightarrow h(U)$ *is a homeomorphism* (the set $h(U)$ being open in $\mathbb{R}^n$).

[For this reason coordinate mappings h considered as mappings onto $h(U)$ are usually called *coordinate homeomorphisms.*]

Consequently, since coordinate neighbourhoods make up an open covering of a manifold $\mathscr{X}$ (this property is enjoyed even by coordinate neighbourhoods which are the domains of definition of the charts of some atlas), *any manifold is a locally Euclidean space, and hence satisfies the first axiom of countability.*

A set $\mathscr{B}$ of open sets of a topological space is said to be its *base* (or, more explicitly, a *base of open sets*) if every open set of that space is a union of sets of $\mathscr{B}$. Spaces having a countable base (i.e. a base containing a countable number of open sets) are said to satisfy the *second axiom of countability* (to be completely separable spaces or spaces satisfying the second axiom of countability).

Clearly, any space that satisfies the second axiom of countability satisfies the first axiom of countability, but the converse is generally wrong.

In these terms Lemma 2 of Lecture 1 states (as applied to the space $\mathbb{R}^n$) that all open rational balls of $\mathbb{R}^n$, i.e. balls whose radii are rational and whose centres have rational coordinates form a base for this space. Thus the *space* $\mathbb{R}^n$ *satisfies the second axiom of countability.*

Nevertheless there are manifolds that do not satisfy the second axiom of countability (so that in this respect the second axiom of countability is in sharp contrast to the first). The simplest example is a disjoint union of a noncountable family of spaces $\mathbb{R}^n$. (A more interesting example will be given in Lecture 11.)

When introducing a topology into the set $\mathscr{X}$, we often only indicate some base for it. It is thus obvious that a *set* $\mathscr{B}$ *of subsets of* $\mathscr{X}$ *may serve as a base of open sets for some topology on* $\mathscr{X}$ *if and only if the intersection of any two sets of* $\mathscr{B}$ *is a union of sets of* $\mathscr{B}$.

In particular, any set of subsets closed under intersections may serve as a base for some topology.

A topological space $\mathscr{X}$ is said to be *Hausdorff* (or *separable space*) if any two of its points p and q have nonintersecting neighbourhoods.

Clearly, any metric space is Hausdorff, and hence, in particular, so is $\mathbb{R}^n$. Nevertheless *there are smooth non-Hausdorff manifolds.*

Example 1. Let $\mathscr{X} = (\mathbb{R}\setminus\{0\}) \cup \{p_0, q_0\}$, where p_0, q_0 are some elements (which are not in $\mathbb{R}\setminus\{0\}$), and let

$$U = \mathscr{X}\setminus\{q_0\} = (\mathbb{R}\setminus\{0\}) \cup \{p_0\},$$

$$V = \mathscr{X}\setminus\{p_0\} = (\mathbb{R}\setminus\{0\}) \cup \{q_0\}.$$

We define bijective mappings

$$h\colon U \to \mathbb{R}, \quad k\colon V \to \mathbb{R}$$

as follows:

$$h(p) = \begin{cases} p & \text{if } p \in \mathbb{R}\setminus\{0\}, \\ 0 & \text{if } p = p_0 \end{cases}$$

for any $p \in U$ and

$$k(q) = \begin{cases} q & \text{if } q \in \mathbb{R} \setminus \{0\}, \\ 0 & \text{if } q = q_0 \end{cases}$$

for any $q \in V$.

Clearly, the charts (U, h) and (V, k) form an atlas and hence define on $\mathscr{X}$ the structure of a smooth (C^ω-) manifold. In the topology of this manifold any two neighbourhoods of p_0 and q_0 intersect, and hence the topology is non-Hausdorff.

The constructed smooth manifold $\mathscr{X}$ is called a *straight line bifurcated at zero* (or *non-Hausdorff straight line with singular point* 0 *of multiplicity* 2).

A *non-Hausdorff straight line with singular point* 0 *of multiplicity* n, where n is an arbitrary integer $\geqslant 2$ is defined similarly.

Very often one has to introduce a smooth manifold structure on a set $\mathscr{X}$ on which there is already a topology, i.e. which is a topological space.

In this case it is always tacitly assumed that the structure must define on $\mathscr{X}$ the given topology, i.e., as it is usual to say, it must be compatible with that topology.

To do this it is necessary that the atlas which specifies on $\mathscr{X}$ a smooth structure having that property should consist of charts (U, h) whose supports U are open and the mappings $h\colon U \to h(U)$ are homeomorphisms. This condition appears to be sufficient.

Proposition 2. Let $\mathbf{A} = \{(U_\alpha, h_\alpha)\}$ *be an atlas on a topological space* $\mathscr{X}$ *such that for any* α *the set* U_α *is open in* $\mathscr{X}$ *and the mapping* $h_\alpha\colon U_\alpha \to h_\alpha(U_\alpha)$ *is a homeomorphism. Then the smoothness defined by* $\mathbf{A}$ *is compatible with the topology of* $\mathscr{X}$.

Proof. According to Proposition 1 a set $O \subset \mathscr{X}$ is open in the topology $\mathbf{T}_\mathbf{A}$ given by the smoothness defined by $\mathbf{A}$ if and only if for any α the set $h_\alpha(O \cap U_\alpha)$ is open (in $\mathbb{R}^n$, and hence also in $h_\alpha(U_\alpha)$), i.e. since mappings h_α are homeomorphisms when the set $O \cap U_\alpha$ is open (in U_α and, therefore, by virtue of U_α being open in $\mathscr{X}$, also open in $\mathscr{X}$).

On the other hand, if the set $O \subset \mathscr{X}$ is open in $\mathscr{X}$, then, since U_α is open in $\mathscr{X}$, the set $O \cap U_\alpha$ is also open in $\mathscr{X}$, and if all the sets $O \cap U_\alpha$ are open in $\mathscr{X}$, then in view of the equation

$$O = \bigcup_\alpha (O \cap U_\alpha) \tag{1}$$

the set O is open in $\mathscr{X}$. Consequently, sets open in the topology $\mathbf{T}_\mathsf{A}$ are precisely the sets open in the topology of $\mathscr{X}$. □

Remark 1. To clarify the proof it is worth noting that in fact it reduces to two trivial remarks. First, the topology of every set U_α is defined solely by the given mapping h_α (the set $O \subset U_\alpha$ is open in U_α if and only if the set $h_\alpha(O)$ is open in $\mathbb{R}^n$); second, for any topological space $\mathscr{X}$ and any its open covering $\{U_\alpha\}$ the topology of $\mathscr{X}$ is uniquely defined (by means of (1)) by the topologies of elements U_α of that covering.

Remark 2. By virtue of Proposition 2 a smooth manifold may be **defined** as a topological space $\mathscr{X}$ on which an atlas is specified, the supports U_α of whose charts (U_α, h_α) are open, and the coordinate mappings h_α are homeomorphisms. The requirement that for any α and β the sets $h_\alpha(U_\alpha \cap U_\beta)$ and $h_\beta(U_\alpha \cap U_\beta)$ should be open in $\mathbb{R}$ will immediately hold. It is this definition that is usually found in the literature (sometimes in a somewhat different wording) in spite of its obvious methodological deficiency (a similar definition of metric spaces, in which this deficiency is most obvious, would state that a metric space is a topological space $\mathscr{X}$ for which a continuous function $\rho: \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ satisfying the common axioms of a metric space is given).

It is in the sets ($\mathbb{R}^n$, $\mathbb{S}^n$, and $\mathbb{R}P^n$) that are already topological (and even metric) spaces that we have introduced smoothnesses in the examples of the preceding lecture. [The distance between the points of $\mathbb{S}^n$ is the angle between the radius vectors and the distance between the points of $\mathbb{R}P^n$, the angle between them as between straight lines in the space $\mathbb{R}^{n+1}$. In both cases (check it!) the axioms of metric spaces hold.]

Problem 1. Prove that the smoothnesses on the spaces $\mathbb{R}^n$, S^n, and $\mathbb{R}P^n$ introduced in Examples 1 to 8 of Lecture 6 are compatible with the topologies of these spaces.

Remark 3. Example 3 of Lecture 6 shows that different (C^ω-) smoothnesses on a topological space compatible with its topology are possible.

When $r = 0$, i.e. for topological manifolds, the situation is quite different.

Let $\mathscr{X}$ be a topological manifold of dimension n, and let (U, h) be a chart in $\mathscr{X}$. This means by definition that $U \subset \mathscr{X}$ and h is a bijective mapping of the form $U \to h(U)$, where $h(U)$ is an open set in $\mathbb{R}^n$. As was repeatedly noted, if (U, h) is a chart of $\mathscr{X}$, then U is open in $\mathscr{X}$ and $h\colon U \to h(U)$ is a homeomorphism. It turns out that when $r = 0$ the converse is also true.

Proposition 3. *A chart (U, h) in a topological manifold $\mathscr{X}$ for which U is open in $\mathscr{X}$ and the mapping $h\colon U \to h(U)$ is a homeomorphism is a chart of $\mathscr{X}$ (is in its maximal atlas $\mathsf{A}_{\max}$).*

To prove this proposition we use the following lemma.

Lemma 1. *If for two charts (U, h) and (V, k) in a topological space $\mathscr{X}$ the sets U and V are open in $\mathscr{X}$ and $h\colon U \to h(U)$ and $k\colon V \to k(V)$ are homeomorphisms, then these charts are compatible (in the C^0-class).*

Proposition 3 is an immediate consequence of Lemma 1, since according to the lemma the chart (U, h) of Proposition 3 is compatible with each chart (V, k) of the maximal atlas $\mathsf{A}_{\max}$ of $\mathscr{X}$ and hence is contained in it. Thus we must only prove Lemma 1.

Proof of Lemma 1. The intersection $U \cap V$ is open in $\mathscr{X}$ (and hence also in U). Consequently, since mappings h and k are homeomorphisms, the sets $h(U \cap V)$ and $k(U \cap V)$ are open in $h(U)$ and $k(V)$ respectively (and hence also in $\mathbb{R}^n$).

In addition, since mappings

$$h|_{U\cap V}\colon U \cap V \to h(U \cap V) \text{ and}$$

$$k|_{U\cap V}\colon U \cap V \to k(U \cap V)$$

are obviously homeomorphisms, their composition

$$h \circ k^{-1}: \quad k(U \cap V) \to h(U \cap V)$$

is also a homeomorphism. The charts (U, h) and (V, k) are thus compatible. □

Proposition 3 implies that for $r = 0$ the charts of $\mathscr{X}$ are characterized purely topologically, i.e. that the *structure of a topological manifold on a topological space $\mathscr{X}$ (when it exists) is uniquely defined*. In other words, the structure of a topological manifold on a topological space $\mathscr{X}$ brings nothing new into this space, and the class of topological manifolds is nothing else but a subclass of the class of all topological spaces.

To find out if a given topological space $\mathscr{X}$ is an n-dimensional manifold, it is necessary to consider all possible open subsets $U \subset \mathscr{X}$ homeomorphic to open subsets of $\mathbb{R}^n$. A space $\mathscr{X}$ is a manifold if and only if charts of the form (U, h), where $h: U \to h(U)$ is some homeomorphism (and $h(U)$ is an open set in $\mathbb{R}^n$), make up an atlas, i.e., since according to Lemma 1 two such charts are compatible, when the sets U are an open covering of $\mathscr{X}$. This proves that *topological manifolds are precisely locally Euclidean spaces*.

Remark 4. There are locally Euclidean spaces (= topological manifolds) into which it is impossible to introduce a smooth (C^r- with $r > 0$) manifold structure. (Such manifolds are called *nonsmoothable* manifolds.) The conditions necessary and sufficient for a smoothness compatible with its topology to exist on a topological manifold are known but quite complicated.

In what follows, unless otherwise stated, topological manifolds are omitted from the consideration.

As already stated in Lecture 6, the case $n = 0$ is not in general excluded. But in fact it is of little interest.

Indeed, when $n = 0$ the space $\mathbb{R}^n$ consists of only one point 0, and hence in a zero-dimensional manifold $\mathscr{X}$ every point has only one coordinate neighbourhood consisting of that point. Thus in a zero-dimensional manifold $\mathscr{X}$ every point (and hence any subset of it) is an open set. A topological space having this property is called *discrete*. Since any discrete space has obviously a unique structure of a zero-dimensional manifold (in which each point is the support of a chart), we therefore see that *zero-dimensional manifolds are precisely discrete spaces.*

Notice that charts of a zero-dimensional manifold do not overlap. We have therefore the right to ascribe to such a manifold an arbitrary class C^r.

Every zero-dimensional manifold (= discrete space) $\mathscr{X}$ is metrizable by a metric in which the distance between any two distinct points is equal to unity. This justifies a visualization of $\mathscr{X}$ as a set scattered into individual isolated points. In what follows the case $n = 0$ will as a rule be excluded from consideration.

A mapping $f\colon \mathscr{X} \to \mathscr{Y}$ of topological spaces is said to be *continuous at a point* $p \in \mathscr{X}$ if for any neighbourhood V of a point $f(p)$ in $\mathscr{Y}$ there is a neighbourhood U of a point p in $\mathscr{X}$ such that $f(U) \subset V$ (cf. the (ε, δ)-definition of a continuous function).

A mapping $f\colon \mathscr{X} \to \mathscr{Y}$ continuous at every point $p \in \mathscr{X}$ is called *continuous on* $\mathscr{X}$.

It is easy to see that a *mapping* $f\colon \mathscr{X} \to \mathscr{Y}$ *is continuous on* $\mathscr{X}$ *if and only if for every open set* $O \subset \mathscr{Y}$ *its complete inverse image* $f^{-1}O = \{p \in \mathscr{X};\ f(p) \in O\}$ *is open in* $\mathscr{X}$. Indeed, if a mapping f is continuous, then for any point $p \in f^{-1}O$ there is, O being a neighbourhood of the point $f(p)$, its neighbourhood U in $\mathscr{X}$ such that $fU \subset O$. But then $U \subset f^{-1}O$ and hence the point p is an interior point of the set $f^{-1}O$. Consequently, the set $f^{-1}O$ is open. Conversely, if for any open set $O \subset \mathscr{Y}$ the set $f^{-1}O \subset \mathscr{X}$ is open, then, in particular, for any point $p \in \mathscr{X}$ and any neighbourhood V of a point $f(p)$ in $\mathscr{Y}$ the set $U = f^{-1}V$ (containing the point p) is open (i.e. it is a neighbourhood of the point p). Since $fU \subset V$, the mapping f is, consequently, continuous at the point p. $\square$

It follows that *homeomorphisms* $f\colon \mathscr{X} \to \mathscr{Y}$ *are precisely continuous bijective mappings for which the inverse mapping* $f^{-1}\colon \mathscr{Y} \to \mathscr{X}$ *is also continuous.*

An injective continuous mapping $f\colon \mathscr{X} \to \mathscr{Y}$ which is a homeomorphism onto a subspace $f\mathscr{X} \subset \mathscr{Y}$ is called a *moneomorphism* (cf. Lecture 1).

Notice that if a set $O \subset \mathscr{X}$ is open in $\mathscr{X}$ and a mapping $f\colon \mathscr{X} \to \mathscr{Y}$ is continuous, then the set fO may not generally be open in $\mathscr{Y}$. For example, this will trivially be the case if the subset $f\mathscr{X}$ of $\mathscr{Y}$ has not a single interior point. (A particular example: $\mathscr{X} = \mathbb{R}$, $\mathscr{Y} = \mathbb{R}^2$ and f is an embedding $x \mapsto (x, 0)$.)

Continuous mappings $f\colon \mathscr{X} \to \mathscr{Y}$ having the property that for any set O open in $\mathscr{X}$ the set $fO \subset \mathscr{Y}$ is open in $\mathscr{Y}$ are called *open*.

Problem 2. Show that:

(a) For every topological space $\mathscr{X}$ the identity mapping id: $\mathscr{X} \to \mathscr{X}$, $p \mapsto p$, is continuous.

(b) For any two continuous mappings of the form $f\colon \mathscr{X} \to \mathscr{Y}$ and $g\colon \mathscr{Y} \to \mathscr{Z}$ their composition

$$g \circ f\colon \mathscr{X} \to \mathscr{Z}, \quad p \mapsto g(f(p))$$

is also a continuous mapping.

Properties (a) and (b) imply that the collection of all topological spaces and all their continuous mappings makes up a *category* of topological spaces denoted by TOP.

If $\mathscr{X}$ and $\mathscr{Y}$ are smooth manifolds (of dimensions n and m, respectively), then for any point $p_0 \in \mathscr{X}$, any continuous mapping $f\colon \mathscr{X} \to \mathscr{Y}$, and any coordinate neighbourhood V of a point $f(p)$ in $\mathscr{Y}$, the neighbourhood U of p_0 in $\mathscr{X}$ for which $fU \subset V$ may also be considered coordinate. Hence if

$$h\colon U \to h(U) \subset \mathbb{R}^n \text{ and } k\colon V \to k(V) \subset \mathbb{R}^m$$

are the coordinate mappings, then

$$\mathring{f} = k \circ (f\,|_U) \circ h^{-1}$$

will define some mapping

$$\mathring{f}\colon h(U) \to k(V) \tag{2}$$

of an open set $h(U)$ of $\mathbb{R}^n$ into an open set $k(V)$ of $\mathbb{R}^m$. This mapping is given by m functions

$$y^j = f^j(x^1, \ldots, x^n), \quad j = 1, \ldots, m, \tag{3}$$

of n variables which express the coordinates $y^1, \ldots, y^m$ of a point $\mathbf{y} = \overset{\circ}{f}(\mathbf{x}) \in k(V) \subset \mathbb{R}^m$ in terms of the coordinates $x^1, \ldots, x^n$ of a point $\mathbf{x} \in h(U) \subset \mathbb{R}^n$ (i.e., in other words, the local coordinates of a point $q = f(p) \in V$ in terms of the local coordinates of a point $p \in U$). We shall say that functions (3) *express* (or *specify*) *a mapping f in charts* (U, h) *and* (V, k) (in local coordinates $x^1, \ldots, x^n$ and $y^1, \ldots, y^m$).

Problem 3. Prove that if (U', h') and (V', k') are other charts having the property that $p_0 \in U'$ and $fU' \subset V'$, then functions f'^j expressing the function f in the charts (U', h') and (V', k') are smooth at a point $h'(p_0)$ if and only if functions (3) are smooth at $h(p_0)$.

In this sense the smoothness property of functions (3) is independent of the choice of charts (U, h) and (V, k).

Definition 3. A continuous mapping $f\colon \mathcal{X} \to \mathcal{Y}$ is called *smooth at a point* $p_0 \in \mathcal{X}$ if in some (and hence in all) charts (U, h) and (V, k) having the property that $p_0 \in U$ and $fU \subset V$, functions (3) expressing the mapping f are smooth functions (of a given class C^r) at $h(p_0)$.

A mapping $f\colon \mathcal{X} \to \mathcal{Y}$ smooth at all points $p \in \mathcal{X}$ is called *smooth*.

Problem 4. Prove that:

(a) For every smooth manifold $\mathcal{X}$ the identity mapping $\mathrm{id}\colon \mathcal{X} \to \mathcal{X}$ is smooth.

(b) For any two smooth mappings of the form $f\colon \mathcal{X} \to \mathcal{Y}$ and $g\colon \mathcal{Y} \to \mathcal{Z}$ their composition

$$g \circ f\colon \mathcal{X} \to \mathcal{Z}, \quad p \mapsto g(f(p))$$

is a smooth mapping.

This means by definition that the *collection of all smooth differentiable manifolds and all their smooth differentiable mappings is a category of differentiable manifolds denoted by* DIFF.

Definition 4. A mapping $f\colon \mathcal{X} \to \mathcal{Y}$ of smooth manifolds is said to be a *diffeomorphism* if

1° it is bijective,

2° smooth,

3° the inverse mapping $f^{-1}\colon \mathcal{Y} \to \mathcal{X}$ is smooth (and hence is also a diffeomorphism).

An obvious example of a diffeomorphism is an arbitrary coordinate mapping $h: U \to h(U)$. (Question: what kind of functions (3) are given by this mapping?)

Conversely, it can easily be seen that *for any open set U of a smooth manifold $\mathscr{X}$ and its diffeomorphism $h: U \to h(U)$ onto an open set $h(U) \subset \mathbb{R}^n$ a pair (U, h) is a chart of $\mathscr{X}$* (it is in its maximal atlas). Indeed, the statement that (U, h) is in a maximal atlas implies that for any chart (V, k) of that atlas a chart (U, h) is compatible with the chart (V, k), i.e. the sets $h(U \cap V)$ and $k(U \cap V)$ are open in $\mathbb{R}^n$ (or equivalently in $h(U)$ and $k(V)$) and the mapping $k \circ h^{-1}: h(U \cap V) \to k(U \cap V)$ is a diffeomorphism. But the first property follows from the fact that the set $U \cap V$ is open in V and in U, and the mappings h and k are homeomorphisms, and the second from the fact that both of these mappings (or more exactly their restrictions to $U \cap V$) are diffeomorphisms (cf. the proof of Proposition 3 and of Lemma 1). □

Manifolds $\mathscr{X}$ and $\mathscr{Y}$ are said to be *diffeomorphisms* if there is at least one diffeomorphism $\mathscr{X} \to \mathscr{Y}$. Such manifolds have the same **differential properties** (properties expressed in terms of smoothnesses) and are in this sense the same. In particular, $\dim \mathscr{X} = \dim \mathscr{Y}$.

Problem 5. Prove that smooth manifolds obtained from a straight line R by introducing the standard smoothness and the smoothness of Example 3, Lecture 6, are diffeomorphic.

Remark 5. It can be shown (try to do it!) that any one-dimensional noncompact smooth manifold which satisfies the second axiom of countability is diffeomorphic to the straight line $\mathbb{R}$ in the standard smoothness (and a compact one to the circle S^1). It is interesting that for $n = 4$ (and only for $n = 4$) there are smoothnesses (very complicated to construct) on $\mathbb{R}^n$ compatible with the topology on $\mathbb{R}^n$, relative to which $\mathbb{R}^n$ is not diffeomorphic to $\mathbb{R}^n$ in the standard smoothness.

Let $\mathscr{X}$ be some set, $\mathscr{Y}$ some smooth manifold, and $f: \mathscr{X} \to \mathscr{Y}$ a bijective mapping. Then *there is a unique smoothness on $\mathscr{X}$ with respect to which f is a diffeomorphism.* All possible pairs of the form $(f^{-1}U, h \circ f)$ are charts of this smoothness where (U, h) is an arbitrary chart of $\mathscr{Y}$. This smoothness is said *to be pulled* back from $\mathscr{Y}$ to $\mathscr{X}$ by means of f.

Clearly, *smoothnesses pulled back from* $\mathscr{Y}$ *to* $\mathscr{X}$ *by means of bijective mappings* $f\colon \mathscr{X} \to \mathscr{Y}$ *and* $g\colon \mathscr{X} \to \mathscr{Y}$ *coincide if and only if the mapping* $g \circ f^{-1}\colon \mathscr{Y} \to \mathscr{Y}$ *is a diffeomorphism.*

Lecture 8

Topological invariance of the dimension of a manifold· The dimension and coverings· Compact spaces· Lebesgue lemma· The upper estimate of the dimension of compact subsets of a space $\mathbb{R}^n$· The monotonicity property of the dimension· Closed sets· The monotonicity of the dimension and closed sets· Direct product of topological spaces· The compactness of the direct product of compact spaces

---

The conclusion made in the preceding lecture that a topological manifold structure on a topological space $\mathscr{X}$ does not bring anything new into the space was in fact somewhat hasty, since a priori it cannot be excluded that the same space $\mathscr{X}$ may have open coverings $\{U_\alpha\}$ and $\{V_\beta\}$ such that every set U_α is homeomorphic to some open set $\overset{\circ}{U}_\alpha$ of $\mathbb{R}^n$, and every set V_β to some open set $\overset{\circ}{V}_\beta$ of $\mathbb{R}^m$, where $n \neq m$, and hence there will exist two different structures of topological manifold on $\mathscr{X}$ in one of which $\mathscr{X}$ is an n-dimensional, and in the other an m-dimensional manifold. The question as to whether or not this may be the case is obviously equivalent to the question as to *whether the dimension is a topological invariant*, i.e. whether homeomorphic manifolds must necessarily have the same dimension.

Suppose such coverings $\{U_\alpha\}$ and $\{V_\beta\}$ exist. Choosing two sets U_{α_0} and V_{β_0} having the property that the set $W = U_{\alpha_0} \cap V_{\beta_0}$ is not empty (clearly, this can always be done) consider in $\mathbb{R}^n$ a set O_1 which is the image of W under the homeomorphism $U_{\alpha_0} \to \overset{\circ}{U}_{\alpha_0}$ and in $\mathbb{R}^m$ a set O_2 which is the image of W under the homeomorphism $V_{\beta_0} \to \overset{\circ}{V}_{\beta_0}$. The sets O_1 and O_2 are open (in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively) and homeomorphic. Conversely, if such sets exist, then they exemplify homeomorphic manifolds of different dimensions.

Thus our question reduces to whether it is possible for open sets in $\mathbb{R}^n$ and $\mathbb{R}^m$ with $n \neq m$ to be homeomorphic. We show that the *answer to this question is negative*, and hence the following **theorem on the topological invariance of the dimension of manifolds** removing all doubts is true.

Theorem 1. *Homeomorphic manifolds have the same dimension.*

To prove Theorem 1 we define an integer dim $\mathscr{X}$ called the dimension for any topological space $\mathscr{X}$. By definition this number is *topologically invariant*, i.e. the same for any homeomorphic spaces. In addition, we prove that for any closed bounded (compact, see below) set $F \subset \mathbb{R}^n$ an inequality

$$\dim F \leqslant n, \tag{1}$$

holds, in the case where F is a cube I^n of $\mathbb{R}^n$ consisting of points $\mathbf{t} = (t^1, \ldots, t^n)$ for which $|t^i| \leqslant 1$ for all $i = 1, \ldots, n$, the equation becoming

$$\dim \mathrm{I}^n = n. \tag{2}$$

This is now sufficient to prove Theorem 1.

Indeed, let O_1 and O_2 be homeomorphic open sets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Since O_1 is an open set, it contains a closed and bounded subset F_1 homeomorphic to the cube I^n. Let F_2 be the image of F_1 under the homeomorphism $O_1 \to O_2$. The subset F_2 of $\mathbb{R}^m$ is also closed and bounded (prove it!). By (1) therefore $\dim F_2 \leqslant m$. On the other hand, since the function dim is topologically invariant, we have $\dim F_2 = \dim F_1 = \dim \mathrm{I}^n = n$. Hence $n \leqslant m$. Since in this reasoning the sets O_1 and O_2 play symmetric roles, by interchanging them we similarly get $m \leqslant n$. Hence $n = m$, which completes the proof of Theorem 1. □

Remark 1. We should bear in mind that for a manifold $\mathscr{X}$ (even smooth) the number dim $\mathscr{X}$ may differ from its dimension in the sense of Definition 5 of Lecture 6 (see Example 4 below). We shall temporarily (only in the present lecture!) denote the dimension of $\mathscr{X}$ in the sense of Definition 5, Lecture 6, by dim′ $\mathscr{X}$. Later (see Remark 3) we shall give this number a topological characterization (and thus prove Theorem 1 once again).

We introduce the function dim by a series of definitions.

Definition 1. The covering $\{U_\alpha\}$ of a space $\mathscr{X}$ is said to be a *refinement* of a covering $\{V_\beta\}$ if for any α there is β such that $U_\alpha \subset V_\beta$.

Definition 2. The covering $\{U_\alpha\}$ of a space $\mathscr{X}$ is said to have *multiplicity* $\leqslant n + 1$ if the intersection of any $(n + 2)$-element subfamily of the covering $\{U_\alpha\}$ is empty. If the covering $\{U_\alpha\}$ has multiplicity $\leqslant n + 1$ but has no multiplicity $\leqslant n$ (i.e. there is in it an $(n + 1)$-element subfamily with nonempty intersection), then the covering $\{U_\alpha\}$ is said to have *multiplicity* $n + 1$.

Definition 3. It is said that dim $\mathscr{X} \leqslant n$ if a finite open covering of multiplicity $\leqslant n + 1$ can be a refinement of any finite open covering of a space $\mathscr{X}$. If dim $\mathscr{X} \leqslant n$ but dim $\mathscr{X} \leqslant n+1$ is false, then it is said that dim $\mathscr{X} = n$.

Example 1. A discrete space is obviously zero-dimensional.

Example 2. The set of all rational numbers (in the topology induced by the topology of the real line) is also zero-dimensional (a finite covering consisting of nonintersecting intervals and having therefore multiplicity 1 can be a refinement of any finite open covering of the set).

Example 3. The set of all irrational numbers is zero-dimensional for similar reasons.

Example 4. Let $\mathscr{X}$ be a non-Hausdorff straight line with singular point O of multiplicity $n + 1$, and let $\{U_\alpha\}$ be a finite open covering of $\mathscr{X}$ such that all $n + 1$ points into which the point O has split are contained in distinct elements of the covering. Then the multiplicity of that covering, as well as that of any open covering which is a refinement of the covering $\{U_\alpha\}$, will be $\geqslant n+1$. This shows that dim $\mathscr{X} \geqslant n$ (whereas dim′ $\mathscr{X} = 1$).

For non-Hausdorff manifolds the invariant dim thus fails to reflect adequately the intuitive idea of dimension.

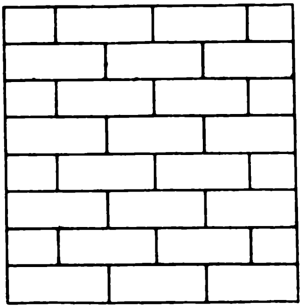
Remark 2. It is not known if dim $\mathscr{X}$ = dim′ $\mathscr{X}$ holds for all Hausdorff manifolds. It can be shown that dim $\mathscr{X} =$

dim′ $\mathscr{X}$ if $\mathscr{X}$ is paracompact (paracompact manifolds will be defined in Lecture 23).

Example 5. It is obvious that a covering consisting of intervals with but slightly overlapping ends (and, consequently, having multiplicity 2) may be a refinement of any finite open covering of the interval $\mathbf{I} = [-1, 1]$. Hence dim $\mathbf{I} \leqslant 1$. Since the equation dim $\mathbf{I} = 0$ is obviously impossible (the interval **I** cannot be covered by a finite family of nonintersecting intervals of length <1), this proves that dim $\mathbf{I} = 1$ (see formula (2)).

Problem 1. Substantiate the above strict reasoning.

Example 6. On covering a square $\mathbf{I}^2 = \{(x, y) \in \mathbb{R}^2, -1 \leqslant x \leqslant 1, -1 \leqslant y \leqslant 1\}$ with rows of bricks so that the joints of bricks in each row are against the middle of the bricks of the neighbouring rows and on enlarging each brick slightly we obtain a finite open covering of the square of multiplicity 3. It can be shown (see below, but we recommend the reader prove this immediately) that this sort of pavement can be a refinement of any finite open covering of the square. Hence dim $\mathbf{I}^2 \leqslant 2$. (As we shall see, the inverse inequality dim $\mathbf{I}^2 \geqslant 2$ is substantially more difficult to prove.)



Our immediate goal is to generalize the construction of Example 6 to an n-dimensional cube

$$\mathbf{I}^n = \{(t_1, \ldots, t_n) \in \mathbb{R}^n, \quad -1 \leqslant t_1 \leqslant 1, \ldots, -1 \leqslant t_n \leqslant 1\}.$$

To do this we shall need some preliminary work.

Definition 4. A topological space $\mathscr{X}$ is said to be *compact* (it is also called *bicompact* in the Russian mathematical literature and *quasi-compact* by Bourbaki) if we can choose a finite subcovering in any of its open coverings $\{U_\alpha\}$.

It is easy to see that *any infinite subset of a compact space has a limiting point* (each neighbourhood of which contains infinitely many points of the subset). Indeed, otherwise each point of the space has a neighbourhood that contains only a finite number of points of the subset. These neighbourhoods make up an open covering with the property that none of its finite subfamilies is a covering (since the total number of points of the subset that are contained in the elements of that subfamily is finite). Because in a compact space such a covering cannot exist, it follows that limiting points do exist. ☐

A subset A of a topological space is said to be *compact* if it is compact in the induced topology (as a subspace).

By an *open covering* in a space $\mathscr{X}$ of the subspace A is meant a family $\{U_\alpha\}$ of open sets of $\mathscr{X}$ such that

$$A \subset \bigcup_\alpha U_\alpha.$$

Then intersections $A \cap U_\alpha$ make up an open covering of A as of a topological space (in the induced topology), and any covering of A can be obtained in this (although not the only) way. Therefore, a *subspace is compact if and only if a finite subcovering can be chosen in any open covering in* $\mathscr{X}$.

The **Heine-Borel theorem** known from calculus states that *any closed bounded subspace of a space* $\mathbb{R}^n$ *is compact. In particular, the cube* $\mathbf{I}^n$ *is compact.*

Recall that the *diameter* of a subset K of a metric space $\mathscr{X}$ is the number

$$d(K) = \sup_{p,q \in K} \rho(p, q),$$

where ρ is a metric on $\mathscr{X}$.

Lemma 1 (Lebesgue). *For any open covering* $\{U_\alpha\}$ *of a compact metric space* $\mathscr{X}$ *there is a positive number* $\varepsilon > 0$

such that any subset K of $\mathscr{X}$ of diameter less than ε is contained in some element of $\{U_\alpha\}$.

Proof. If there is no such ε, then for any $n > 0$ a subset K_n of the diameter $< 1/n$ can be found in $\mathscr{X}$ that is not contained in any element of $\{U_\alpha\}$. Choosing in K_n an arbitrary point p_n, consider the set $\{p_n\}$. Since $\mathscr{X}$ is compact, this set has at least one limiting point p_0. Let $p_0 \in U_{\alpha_0}$, and let d be the distance from p_0 to $\mathscr{X} \setminus U_{\alpha_0}$ (i.e. $d = \inf \rho(p_0, p)$, where $p \in \mathscr{X} \setminus U_\alpha$). If $n > 2/d$ and $\rho(p_0, p_n) < d/2$, then

$$\rho(p_0, p) \leqslant \rho(p_0, p_n) + \rho(p_n, p) \leqslant \frac{d}{2} + \frac{1}{n} < d$$

for any point $p \in K_n$, and hence, despite the assumption, $K_n \subset U_{\alpha_0}$. □

The number $\varepsilon > 0$ of the Lebesgue lemma is called the *Lebesgue number* of a covering $\{U_\alpha\}$.

The covering $\{V_\beta\}$ of a metric space $\mathscr{X}$ is said to be a *ε-covering* if $d(V_\beta) < \varepsilon$ for any β.

Corollary 1. *If ε is the Lebesgue number of a covering $\{U_\alpha\}$, then any finite open ε-covering $\{V_\beta\}$ of a compact metric space $\mathscr{X}$ is a refinement of $\{U_\alpha\}$.* □

Corollary 2. *For a compact metric space $\mathscr{X}$, the inequality* $\dim \mathscr{X} \leqslant n$ *holds if and only if for any $\varepsilon > 0$ there is a finite open ε-covering of $\mathscr{X}$ of multiplicity $\leqslant n + 1$.* □

Now we are in a position to extend Example 6 to any n (and to any closed bounded set $F \subset \mathbb{R}^n$).

Proposition 1. *For every closed bounded set $F \subset \mathbb{R}^n$ an inequality*

$$\dim F \leqslant n$$

holds.

Proof. Let $N > 0$. By induction on n construct for any $n \geqslant 1$ some special covering of $\mathbb{R}^n$ consisting of closed sets and call it *a pavement of rank N*. Points of $\mathbb{R}^n$, $n \geqslant 2$, will be identified with pairs $(\mathbf{t}, t)$, where $\mathbf{t} \in \mathbb{R}^{n-1}$ and $t \in \mathbb{R}$.

By definition a pavement of rank N of a straight line $\mathbb{R}$ consists of the intervals $[(a - 1)/N, a/N]$, $-\infty < a < +\infty$.

Let a pavement of rank N of a space $\mathbb{R}^{n-1}$, $n \geqslant 2$, be already constructed. As a pavement of rank N of $\mathbb{R}^n$ we take the covering which consists of all sets of the form

$$K \times [(a-1)/N,\ a/N]$$
$$= \{(\mathbf{t},\ t) \in \mathbb{R}^n,\ \mathbf{t} \in K,\ (a-1)/N \leqslant t \leqslant a/N\},$$

where a is an arbitrary integer and K, with a odd, is an element of the pavement of rank N of $\mathbb{R}^{n-1}$ and, with a even an element of the pavement of rank N of $\mathbb{R}^{n-1}$ shifted by a vector $(1/2N, \ldots, 1/2N)$.

By induction on n we immediately prove that:

(a) the pavement of rank N of $\mathbb{R}^n$ is a covering of $\mathbb{R}^n$ that consists of cubes with side $1/N$,

(b) any point of $\mathbb{R}^n$ is in no more than $n+1$ cubes of a pavement of rank N, with a point being precisely in $n+1$ cubes if and only if it is of the form

$$\left(\frac{a_1}{2N}, \ldots, \frac{a_{n-1}}{2N}, \frac{a_n}{N}\right),$$

where $a_1, \ldots, a_{n-1}, a_n$ are integers.

Since the diameter of a cube with side a in $\mathbb{R}^n$ is obviously equal to $a\sqrt{n}$, we therefore see that a pavement of rank N is a *closed* (consisting of closed sets) ($\sqrt{n}/N$)-covering of $\mathbb{R}^n$ of multiplicity $n+1$. To obtain an open covering we choose some number $\delta > 0$ and replace each cube of the pavement by its δ-neighbourhood (obviously having a diameter of $2\delta + (\sqrt{n}/N)$). With δ sufficiently small (namely, $\delta < 1/2N$), this does not affect the multiplicity of the covering. Since for any $\varepsilon > 0$ there are N and δ such that $\varepsilon > 2\delta + (\sqrt{n}/N)$, we therefore see that *for any $\varepsilon > 0$ there is an open ε-covering of $\mathbb{R}^n$ of multiplicity $n+1$.*

Any closed bounded (i.e. compact) subset F of $\mathbb{R}^n$ intersects only a finite number of elements of that covering, and hence that covering cuts out on F a finite open ε-covering of multiplicity $\leqslant n+1$. Therefore $\dim F \leqslant n$. □

Thus, of the two formulas (1) and (2) we need for proving Theorem 1 we have already proved (1). To complete the proof of Theorem 1 we should only prove the inequality $\dim \mathbf{I}^n \geqslant n$ (which in fact is the most difficult part of the proof). In order not to interrupt the presentation, we shall complete the proof in the next lecture, proceeding for the time being to the discussion of formula (1) in more detail.

Better use of the pavement construction will show that (1) is true for any subset F of $\mathbb{R}^n$ (not necessarily closed or bounded).

Problem 2. Prove (1) for arbitrary subsets $F \subset \mathbb{R}^n$.

In particular, for $F = \mathbb{R}^n$ we get

$$\dim \mathbb{R}^n \leqslant n. \tag{3}$$

Therefore (1) is a consequence of a formally more strict inequality

$$\dim F \leqslant \dim \mathbb{R}^n \tag{4}$$

(in fact, as we shall later see, there is an equality sign in formula (3):

$$\dim \mathbb{R}^n = n, \tag{5}$$

and so (1) and (4) are equivalent).

Formula (4) states that for a topological space $\mathbb{R}^n$, the dimension of its arbitrary subspace is at most the dimension of the space itself, which agrees well with our geometric intuition. It is natural to expect that a similar **property of monotonicity of dimension**

$$\dim F \leqslant \dim \mathscr{X} \tag{6}$$

is valid also for a subspace F of an arbitrary topological space $\mathscr{X}$. As we shall see from an example of Lecture 10, however, *inequality* (6) *is generally false*, and for it to be true it is necessary to impose some additional conditions on F or on $\mathscr{X}$.

Definition 5. A subset F of a topological space $\mathscr{X}$ is said to be *closed* if its complement is open.

For subsets of metric spaces (and, in particular, for subsets of $\mathbb{R}^n$) this definition is known from calculus (and we have already used it for these spaces).

Problem 3. Prove that *in any manifold $\mathscr{X}$ every one-point set is closed* (i.e., as is customary to say, points in $\mathscr{X}$ are closed).

Problem 4. Prove that *every closed subset of a compact space is compact.* [Hint. Make use of the fact that adding an open set $\mathscr{X} \setminus F$ to an arbitrary open covering of a set F yields an open covering of the entire space $\mathscr{X}$.]

It is clear that:

1°. An empty set $\varnothing$ and the entire space $\mathscr{X}$ are closed.

2° The intersection of any family of closed sets is closed.

3° The union of any finite family of closed sets is closed.

Cf. the corresponding properties 1° to 3° of open sets in Lecture 7.

It follows directly from Property 2° of closed sets that for *any subset A of a topological space there is the smallest closed set containing A* (it is the intersection of all closed sets containing A). This closed set is called the *closure* of the set A and is denoted by Cl A in the English topological literature and by $\bar{A}$ or $[A]$ in the Russian literature.

Clearly, *a point $p \in \mathscr{X}$ is in $\bar{A}$ if and only if any neighbourhood of the point intersects with A.* This means that

$$\mathscr{X} \setminus \bar{A} = \operatorname{Int} (\mathscr{X} \setminus A).$$

[Notice that for Int A there is a dual formula $\mathscr{X} \setminus \operatorname{Int} A = \overline{\mathscr{X} \setminus A}$.]

Problem 5. Prove that the following properties of $f\colon \mathscr{X} \to \mathscr{Y}$ are equivalent:

(a) The mapping f is continuous.

(b) For any closed set $C \subset \mathscr{Y}$ the set $f^{-1}C \subset \mathscr{X}$ is closed.

(c) For any set $A \subset \mathscr{X}$ an inclusion

$$f(\bar{A}) \subset \overline{f(A)}$$

holds.

(d) For any set $B \subset \mathscr{Y}$ an inclusion

$$\overline{f^{-1}(B)} \subset f^{-1}(\bar{B})$$

holds.

Remark 3. We have seen that the invariant dim is not always adequate to an intuitive idea of dimension. We may try to set the things straight by introducing a new invariant dim′. For any topological space $\mathscr{X}$ it will be assumed that $\dim' \mathscr{X} \leqslant n$ if for an arbitrary neighbourhood U of every point $p \in \mathscr{X}$ there is such a neighbourhood V of the point contained in U, that for its closure $\bar{V}$ in U the inequality $\dim \bar{V} \leqslant n$ holds. (The restriction "in U" is essential here since the closure of V in $\mathscr{X}$ may be substantially greater). If $\dim' \mathscr{X} \leqslant n$ but $\dim' \mathscr{X} \leqslant n-1$ is false, then by definition $\dim' \mathscr{X} = n$. Since any point of every open set $U \subset \mathbb{R}^n$ has a neighbourhood whose closure is homeomorphic to the cube $\mathbf{I}^n$, for any manifold $\mathscr{X}$ (here we use the yet-unproved formula (2)) the number $\dim' \mathscr{X}$ coincides with the dimension of the manifold in the sense of Definition 5 of Lecture 6 (see Remark 1 above). The conditions necessary and sufficient for the equation

$$\dim \mathscr{X} = \dim' \mathscr{X}$$

to hold are still unknown (see Example 4 and Remark 2 above).

As already stated, the dimension of a subspace may be greater than that of the entire space.

This is impossible for closed subspaces, however.

Proposition 2. *For any closed subspace F of an arbitrary topological space $\mathscr{X}$*

$$\dim F \leqslant \dim \mathscr{X}. \tag{7}$$

Proof. Let $\{U_\alpha\}$ be an arbitrary finite open covering of a subspace F. The statement that U_α is open in F implies that there are open sets U'_α in $\mathscr{X}$ such that $U_\alpha = U'_\alpha \cap F$ for any α. The sets U'_α together with $\mathscr{X} \setminus F$ form a finite open covering of $\mathscr{X}$. Therefore there is a finite open covering $\{V_\beta\}$ of multiplicity $\leqslant n+1$, where $n = \dim \mathscr{X}$, which is a refinement of the covering of $\mathscr{X}$. Consider all nonempty sets of the form $V'_\beta \cap F$. These sets clearly form a finite open covering of F of multiplicity $\leqslant n+1$, which is a refinement of $\{U_\alpha\}$. Thus a finite open covering of multiplicity $\leqslant n+1$ may be a refinement of every finite open covering of subspace F. Hence $\dim F \leqslant n$. □

In an example to be constructed in Lecture 10 the subspace F will be open (and the space $\mathscr{X}$ Hausdorff and compact).

It can be shown (it is a difficult theorem) that for inequality (6) to hold for any subspace F it is sufficient that $\mathscr{X}$ be a metrizable space and satisfy the second axiom of countability (generally, metrizable spaces that satisfy the second axiom of countability form in a certain sense the most natural field for constructing a dimension theory, outside this class of spaces almost any "natural" property of dimension being false). In the example of Lecture 10 the space $\mathscr{X}$ is therefore either not metrizable or fails to satisfy the second axiom of countability. (In fact, as can easily be shown, it is both nonmetrizable and does not satisfy the second axiom of countability. This does not prevent it from being Hausdorff and compact.)

Remark 4. From Proposition 2 it follows, in particular, that $\dim \mathrm{I}^n \leqslant \dim \mathbb{R}^n$. By virtue of formula (2) to be proved in Lecture 9 this means that $n \leqslant \dim \mathbb{R}^n$. Together with formula (3), which has not been proved yet, this yields equation (5).

Along with the approach outlined above another, more conceptual, approach to the proof of formula (3) is possible.

It rests on one simple but important topological construction, which, in special cases, has already been presented in the course of calculus.

Let $\mathscr{X}$ and $\mathscr{Y}$ be arbitrary topological spaces, and let $\mathscr{X} \times \mathscr{Y}$ be the set of all pairs (p, q), where $p \in \mathscr{X}$ and $q \in \mathscr{Y}$. If $U \subset \mathscr{X}$ and $V \subset \mathscr{Y}$, then $U \times V \subset \mathscr{X} \times \mathscr{Y}$, and it is clear that the set of all subsets of the form $U \times V$, where U is open in $\mathscr{X}$ and V is open in $\mathscr{Y}$, is closed under intersections. This set is therefore a base for some topology on $\mathscr{X} \times \mathscr{Y}$.

Definition 6. The resulting topological space is called the *direct* (or *Cartesian*) *product* of spaces $\mathscr{X}$ and $\mathscr{Y}$. It is denoted by the same symbol $\mathscr{X} \times \mathscr{Y}$.

It is easy to see that the *direct product $\mathscr{X} \times \mathscr{Y}$ of Hausdorff spaces $\mathscr{X}$ and $\mathscr{Y}$ is Hausdorff*. Indeed, let (p_1, q_1) and (p_2, q_2) be two distinct points of the space $\mathscr{X} \times \mathscr{Y}$.

Then either $p_1 \neq p_2$ or $q_1 \neq q_2$. Let $p_1 \neq p_2$ for definiteness. Since under the hypothesis $\mathscr{X}$ is a Hausdorff space, the points p_1 and p_2 have nonintersecting neighbourhoods U_1 and U_2 in $\mathscr{X}$. Then the sets $U_1 \times \mathscr{Y}$ and $U_2 \times \mathscr{Y}$ are nonintersecting neighbourhoods of the points (p_1, q_1) and (p_2, q_2) in the product $\mathscr{X} \times \mathscr{Y}$. □

Similarly defined is the direct product $\mathscr{X}_1 \times \ldots \times \mathscr{X}_n$ of any finite family of topological spaces. It is Hausdorff if so are the spaces $\mathscr{X}_1, \ldots, \mathscr{X}_n$.

Problem 6. Prove that

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n \text{ times}}.$$

[Hint. For any ball of a space $\mathbb{R}$, there are an inscribed and a described cube with faces parallel to the coordinate planes.]

For the sets $U \times V$ to make up the base for the space $\mathscr{X} \times \mathscr{Y}$ there is no need for U and V to range over all the open sets of spaces $\mathscr{X}$ and $\mathscr{Y}$, respectively. It is obviously sufficient for U's to range over some base for $\mathscr{X}$ and for V's to range over some base for $\mathscr{Y}$. Therefore, *the direct product of spaces that satisfy the second axiom of countability also satisfies this axiom.*

Again we see, in particular, that the space $\mathbb{R}^n$ satisfies the second axiom of countability.

Consideration of the simplest examples (say cubes I^n) suggests that for any two spaces $\mathscr{X}$ and $\mathscr{Y}$ the equation

$$\dim (\mathscr{X} \times \mathscr{Y}) = \dim \mathscr{X} + \dim \mathscr{Y} \tag{8}$$

should hold. However, examples, which we cannot present here because of their complexity, show that *equation* (8) *is false even for compact metric spaces.*

Problem 7. Prove that any compact metric space satisfies the second axiom of countability.

Moreover, for arbitrary topological spaces $\mathscr{X}$ and $\mathscr{Y}$ even

$$\dim (\mathscr{X} \times \mathscr{Y}) \leqslant \dim \mathscr{X} + \dim \mathscr{Y} \tag{9}$$

may fail. Nevertheless, it can be shown that *in the class of metric spaces satisfying the second axiom of countability* (and in the class of Hausdorff compact spaces) *inequality* (9) *is true*. Since, as is easy to see, dim $\mathbb{R} = 1$ (cf. Example 5 above), this proves inequality (3) again. The proof of formula (9) for metric spaces satisfying the second axiom of countability is rather complicated and there is no room for presenting it here.

The statement that formula (9) is true in the class of Hausdorff compact spaces implies, in particular, that the following proposition is true.

Proposition 3. *The direct product of compact topological spaces is compact.*

Before proving the proposition we shall make some remarks on the projection

$$(10) \qquad \mathscr{X} \times \mathscr{Y} \to \mathscr{X}, \quad (p, q) \mapsto p.$$

The inverse image of projection (10) of an arbitrary open set $U \subset \mathscr{X}$ is an (open by definition) set $U \times \mathscr{Y}$. Hence *projection* (10) *is a continuous mapping*.

Moreover, since every set of the form $U \times V$ is projected onto a set U, *projection* (10) *maps every open set of the space* $\mathscr{X} \times \mathscr{Y}$, which is an open mapping, *onto an open set of the space* $\mathscr{X}$.

Similar statements are, of course, also true for the projection $\mathscr{X} \times \mathscr{Y} \to \mathscr{Y}$, $(p, q) \mapsto q$, as well as for the projections $\mathscr{X}_1 \times \ldots \times \mathscr{X}_n \to \mathscr{X}_k$ of a product of any number of spaces on each of the factors.

At the same time projection (10) may transform a closed set into a nonclosed one (under projection $\mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x$, for example, the hyperbola $xy = 1$, which is a closed set of the plane $\mathbb{R}^2$, is transformed into a nonclosed set $\mathbb{R}\setminus\{0\}$ of the axis $\mathbb{R}$). But *if the space* $\mathscr{Y}$ *is compact, then projection* (10) *maps every closed set of space* $\mathscr{X}$ *onto a closed set of* $\mathscr{X}$, which is a *closed mapping*. Suppose a point $p_0 \in \mathscr{X}$ is not in the projection pr $F \subset \mathscr{X}$ of a closed set $F \subset \mathscr{X} \times \mathscr{Y}$, i.e. suppose $\{p_0\} \times \mathscr{Y} \subset (\mathscr{X} \times \mathscr{Y})\setminus F$. Since $(\mathscr{X} \times \mathscr{Y})\setminus F$ is an open set and hence any its point is an interior point, it follows that for every point $q \in \mathscr{Y}$ there is a neighbourhood of that point $V_q \subset \mathscr{Y}$ and a neighbourhood $U(q) \subset \mathscr{X}$ of the point p_0 such that $U(q) \times V_q \subset (\mathscr{X} \times \mathscr{Y})\setminus F$. All neighbourhoods V_q, $q \in \mathcal{Y}$, make up an open covering of the space $\mathcal{Y}$ and therefore, $\mathcal{Y}$ being a compact space, it is possible to choose from that covering a finite covering $V_{q_1}, \ldots, V_{q_n}$. Let

$$U = U(q_1) \cap \ldots \cap U(q_n).$$

The set U is open, contains the point p_0 and has the property that for any point $p \in U$ all the points of the product $\mathcal{X} \times \mathcal{Y}$ of the form (p, q), $q \in \mathcal{Y}$, are not in F (if $q \in V_{q_i}$, then $(p, q) \in U(q_i) \times V_{q_i} \subset (\mathcal{X} \times \mathcal{Y}) \setminus F$). This means that $\{p\} \times \mathcal{Y} \subset (\mathcal{X} \times \mathcal{Y}) \setminus F$, i.e. that $p \notin \operatorname{pr} F$. This proves that every point $p_0 \notin \operatorname{pr} F$ has a neighbourhood $U \subset \mathcal{X}$ such that $U \subset \mathcal{X} \setminus \operatorname{pr} F$, i.e. the point p_0 is an interior point of the set $\mathcal{X} \setminus \operatorname{pr} F$. Hence this set is open, and therefore the set $\operatorname{pr} F$ is closed. □

Now we are in a position to prove Proposition 3.

Proof of Proposition 3. Let $\mathcal{X}$ and $\mathcal{Y}$ be compact spaces, and let $\{W_\alpha\}$ be an arbitrary open covering of the space $\mathcal{X} \times \mathcal{Y}$. An open set $O \subset \mathcal{X}$ will be said (only in this proof!) to be *distinguished* if the subset $O \times \mathcal{Y}$ of the product $\mathcal{X} \times \mathcal{Y}$ is contained in the union of a finite subfamily of $\{W_\alpha\}$. For any point $p \in \mathcal{X}$ the subspace $\{p\} \times \mathcal{Y}$ of the product $\mathcal{X} \times \mathcal{Y}$ is homeomorphic (prove it!) to the space $\mathcal{Y}$ and hence compact. It can therefore be covered by a finite subfamily of $\{W_\alpha\}$. Let F be a closed subset of the space $\mathcal{X} \times \mathcal{Y}$ which is in $\mathcal{X} \times \mathcal{Y}$ a complement of the union G of all elements of that subfamily. Since under the hypothesis $\mathcal{Y}$ is compact, it follows from what has been proved that the projection $\operatorname{pr} F$ of the set F is closed in $\mathcal{X}$ and hence its complement O is open. On the other hand, by construction the set $O \times \mathcal{Y}$ is contained in G. Hence the set O is distinguished. Since $p \in O$, this proves that distinguished sets make up a covering of $\mathcal{X}$. The space $\mathcal{X}$ being compact, it is possible to choose a finite subcovering from that covering, i.e. to find a finite family of distinguished sets that cover $\mathcal{X}$. But it is clear that the union of any finite family of distinguished sets is distinguished. Hence $\mathcal{X}$ is distinguished, and therefore the covering $\{W_\alpha\}$ contains a finite subcovering. □

The proof of inequality (9) for Hausdorff compact spaces is also too complicated to be presented here.

Lecture 9

The drum theorem· The Brouwer fixed-point theorem· Cube separation theorem· Normal and completely normal spaces· Separation extension· The Lebesgue theorem on coverings of a cube· The lower estimate of the dimension of a cube

---

In the preceding lecture we have proved that dim $\mathbf{I}^m \leqslant n$. The main goal of this lecture is to prove the converse and thus to establish that

$$\dim \mathbf{I}^n = n \quad \text{for any} \quad n \geqslant 0.$$

To do this we must begin from quite afar.

Definition 1. A subspace A of a topological space $\mathscr{X}$ is said to be a *retract* of $\mathscr{X}$ if there is a continuous mapping

$$r\colon \mathscr{X} \to A$$

(called a *retraction*) such that $r(a) = a$ for any point $a \in A$.

Recall that by B^n we denote a *unit ball* of $\mathbb{R}^n$ that consists of points $\mathbf{x} \in \mathbb{R}^n$ for which $|\mathbf{x}| \leqslant 1$ (where as ever $|\mathbf{x}| = \sqrt{x_1^2 + \ldots + x_n^2}$ if $\mathbf{x} = (x_1, \ldots, x_n)$) and by S^{n-1} we denote a *unit sphere* (a subset of B^n that consists of points $\mathbf{x} \in \mathbb{R}^n$ for which $|\mathbf{x}| = 1$).

Theorem 1. *The sphere S^{n-1} is not a retract of the ball B^n.*

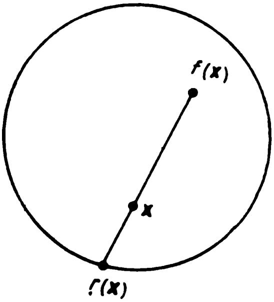
For $n = 2$ this theorem theoretically explains why it is possible to pull a film over a circle, i.e. to make a drum. That is why Theorem 1 is usually called the **drum theorem**.

In spite of the intuitive obviousness of the theorem it is rather difficult to prove. Its proof requires a wealth of new ideas. We therefore put it off until Lecture 24.

An obvious corollary of Theorem 1 is the following **Brouwer theorem.**

Theorem 2. *For any continuous mapping* $f\colon B^n \to B^n$ *of a ball* B^n *into itself there is a point* $\mathbf{x} \in B^n$ *such that* $f(\mathbf{x}) = \mathbf{x}$.

Proof. If $f(\mathbf{x}) \neq \mathbf{x}$, then a straight line passing through points $\mathbf{x}$ and $f(\mathbf{x})$ is defined. Let $r(\mathbf{x})$ be that point of the two points of intersection of the straight line with the sphere S^{n-1} which is not separated from the point $\mathbf{x}$ by the point $f(\mathbf{x})$. If $f(\mathbf{x}) \neq \mathbf{x}$ for all points $\mathbf{x}$, then this construction defines (obviously a continuous) mapping $r\colon B^n \to S^{n-1}$. If $\mathbf{x} \in S^{n-1}$, then the point $\mathbf{x}$ itself is not separated from the point $\mathbf{x}$ by the point of intersection. Therefore $r(\mathbf{x}) = \mathbf{x}$, i.e. r is a retraction. Since the existence of such a mapping is contrary to Theorem 1, the inequality $f(\mathbf{x}) \neq \mathbf{x}$ cannot be satisfied for all points $\mathbf{x} \in B^n$. □

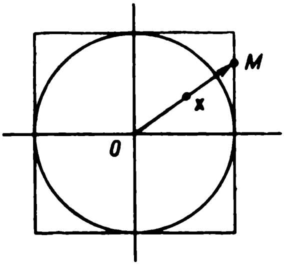


Corollary 1. *For any continuous mapping* $f\colon I^n \to I^n$ *of a cube* I^n *into itself there is a point* $\mathbf{t}_0 \in I^n$ *such that* $f(\mathbf{t}_0) = \mathbf{t}_0$.

Proof. It suffices to notice that the *cube* I^n *is homeomorphic to the ball* B^n. [The homeomorphism $B^n \to I^n$ can be given as follows

$$\mathbf{x} \mapsto \lambda(\mathbf{x})\,\mathbf{x}, \quad \mathbf{x} \in B^n,$$

where $\lambda(\mathbf{0}) = 0$ and $\lambda(\mathbf{x})$ for $\mathbf{x} \neq 0$ is the length of vector $\mathbf{OM}$ collinear with the vector $\mathbf{x}$ and such that the point M



is on the boundary of the cube $\mathbf{I}^n$ (it is easy to see that the vector $\mathbf{OM}$ is uniquely defined up to a sign by this condition).] □

Remark 1. Theorem 1 is easy to derive from Theorem 2. Indeed, if the retraction $r\colon \mathsf{B}^n \to \mathsf{S}^{n-1}$ existed, then the composite mapping

$$i \circ a \circ r\colon \quad \mathsf{B}^n \to \mathsf{B}^n,$$

where $a\colon \mathsf{S}^{n-1} \to \mathsf{S}^{n-1}$ is an *antipodal mapping* $\mathbf{x} \mapsto -\mathbf{x}$ and i is an embedding $\mathsf{S}^{n-1} \to \mathsf{B}^n$, would be a mapping $\mathsf{B}^n \to \mathsf{B}^n$ without fixed points.

Definition 2. Let A and B be disjoint subsets of a topological space $\mathscr{X}$. A closed set $C \subset \mathscr{X}$ is said to *separate* A and B if its complement $\mathscr{X} \setminus C$ is a union of disjoint (and automatically, open) sets U and V such that $A \subset U$ and $B \subset V$.

Of course, a partition C may be an empty set (this will be the case if the space $\mathscr{X}$ itself is a union of open nonintersecting sets U and V that contain sets A and B respectively).

For any $k = 1, \ldots, n$ and any $\varepsilon = \pm 1$ we use the symbol $\mathbf{I}_k^{n-1}(\varepsilon)$ to denote a face of a cube $\mathbf{I}^n$ consisting of points $\mathbf{t} = (t_1, \ldots, t_n) \in \mathbf{I}^n$ for which $t_k = \varepsilon$.

Theorem 3 (cube separation). *The intersection* $C_1 \cap \ldots \cap C_n$ *of any sets* $C_1, \ldots, C_n$ *separating in a cube* $\mathbf{I}^n$ *faces* $\mathbf{I}_1^{n-1}(-1), \ldots, \mathbf{I}_n^{n-1}(-1)$ *and* $\mathbf{I}_n^{n-1}(+1), \ldots, \mathbf{I}_n^{n-1}(+1)$ *is not empty*:

$$C_1 \cap \ldots \cap C_n \neq \varnothing.$$

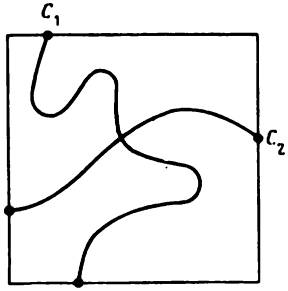
Proof. Under the hypothesis, for any $k = 1, \ldots, n$

$$\mathbf{I}^n \setminus C_k = U_k(-1) \cup U_k(+1),$$

where $U_k(-1)$ and $U_k(+1)$ are nonintersecting open sets such that $\mathbf{I}_k^{n-1}(-1) \subset U_k(-1)$ and $\mathbf{I}_k^{n-1}(+1) \subset U_k(+1)$. For every point $\mathbf{t} \in \mathbf{I}^n$ we put

$$\rho_k(\mathbf{t}) = \inf_{\mathbf{s} \in C_k} |\mathbf{t} - \mathbf{s}|, \quad k = 1, \ldots, n.$$

Notice that in view of the set C_k being compact the equation $\rho_k(\mathbf{t}) = 0$ is equivalent to the inclusion $\mathbf{t} \in C_k$.



$C_1 \cap C_2 \neq \varnothing$

Lemma 1. *The number* $\rho_k(\mathbf{t})$ *has the property that* $|t_k - \varepsilon\rho_k(\mathbf{t})| \leqslant 1$ *if* $\mathbf{t} = (t_1, \ldots, t_n) \in U_k(\varepsilon)$.

It follows from the lemma that the formula

$$f(\mathbf{t}) = (t_1 - \tau_1, \ldots, t_n - \tau_n),$$

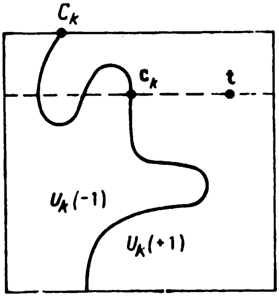
where $$\tau_k = \begin{cases} \varepsilon\rho_k(t) & \text{if } \mathbf{t} \in U_k(\varepsilon), \\ 0 & \text{if } \mathbf{t} \in C_k, \end{cases} \quad k = 1, \ldots, n,$$

correctly defines some (obviously continuous) mapping $f: \mathbf{I}^n \to \mathbf{I}^n$ having the property that $f(\mathbf{t}) = \mathbf{t}$ if and only if $\rho_k(\mathbf{t}) = 0$ for any $k = 1, \ldots, n$, i.e. if $\mathbf{t} \in C_1 \cap \ldots \cap C_n$. Since, however, according to Corollary 1 of Theorem 2 there is a point $\mathbf{t}_0$ in the cube $\mathbf{I}^n$ for which $f(\mathbf{t}_0) = \mathbf{t}_0$, this proves that $C_1 \cap \ldots \cap C_n \neq \varnothing$. □

It remains to prove Lemma 1.

Proof of Lemma 1. Let a point $\mathbf{t} \in \mathbf{I}^n$ be in an open set $U_k(\varepsilon)$. Since the set C_k separates faces $\mathbf{I}_k^{n-1}(-1)$ and

$\mathbf{I}_k^{n-1}(+1)$, the perpendicular to these faces passing through $\mathbf{t}$ intersects C_k at some point $\mathbf{c}_k$ which is between the point $\mathbf{t}$ and its projection onto the face $\mathbf{I}_k^{n-1}(-\varepsilon)$. If $\varepsilon = +1$,



then for the kth coordinate s_k of $\mathbf{c}_k$ we have $s_k \leqslant t_k$, and if $\varepsilon = -1$ we have $s_k \geqslant t_k$, with $|s_k - t_k| = \sigma_k$, where σ_k is the distance between the points $\mathbf{t}$ and $\mathbf{c}_k$. Hence

$$s_k = t_k - \varepsilon\sigma_k.$$

Since by definition

$$\sigma_k \leqslant \rho_k(\mathbf{t}),$$

it follows that for $\varepsilon = +1$,

$$\begin{aligned} t_k - \varepsilon\rho_k(\mathbf{t}) &= t_k - \rho_k(\mathbf{t}) \geqslant t_k - \sigma_k = t_k - \varepsilon\sigma_k \\ &= s_k \geqslant -1 \end{aligned}$$

and for $\varepsilon = -1$,

$$\begin{aligned} t_k - \varepsilon\rho_k(\mathbf{t}) &= t_k + \rho_k(\mathbf{t}) \leqslant t_k + \sigma_k = t_k - \varepsilon\sigma_k \\ &= s_k \leqslant 1. \end{aligned}$$

On the other hand,

$$t_k - \varepsilon\rho_k(\mathbf{t}) = t_k - \rho_k(\mathbf{t}) \leqslant t_k \leqslant 1$$

for $\varepsilon = +1$ and

$$t_k - \varepsilon\rho_k(\mathbf{t}) = t_k + \rho_k(\mathbf{t}) \geqslant t_k \geqslant -1$$

for $\varepsilon = -1$. Thus in all cases

$$|t_k - \varepsilon\rho_k(\mathbf{t})| \leqslant 1. \ \square$$

Remark 2. Theorem 1 can also be easily derived from Theorem 3. Indeed, let $\dot{\mathbf{I}}^n$ be a *boundary* of a cube $\mathbf{I}^n$

(that consists of points $\mathbf{t} \in \mathbf{I}^n$ for which $t_k = \pm 1$ at least for one $k = 1, \ldots, n$), and let S_k, $k = 1, \ldots, n$, be its subset consisting of points $\mathbf{t} \in \dot{\mathbf{I}}^n$ for which $t_k = 0$. Clearly S_k is a set in $\dot{\mathbf{I}}^n$ separating faces $\mathbf{I}_k^{n-1}(-1)$ and $\mathbf{I}_k^{n-1}(+1)$. For any retraction $r\colon \mathbf{I}^n \to \dot{\mathbf{I}}^n$ the inverse image $C_k = r^{-1}S_k$ of the set S_k separates the inverse images $r^{-1}\mathbf{I}_k^{n-1}(-1)$ and $r^{-1}\mathbf{I}_k^{n-1}(+1)$ of faces $\mathbf{I}_k^{n-1}(-1)$ and $\mathbf{I}_k^{n-1}(+1)$ and hence in view of the inclusions $\mathbf{I}_k^{n-1}(\varepsilon) \subset r^{-1}\mathbf{I}_k^{n-1}(\varepsilon)$, the faces $\mathbf{I}_k^{n-1}(-1)$ and $\mathbf{I}_k^{n-1}(+1)$ themselves. On the other hand, since it is obvious that $S_1 \cap \ldots \cap S_n = \varnothing$, it is also obvious that $C_1 \cap \ldots \cap C_n = \varnothing$, which is impossible by virtue of Theorem 3. Hence for the pair $(\mathbf{I}^n, \dot{\mathbf{I}}^n)$, and therefore for the homeomorphic pair $(\mathbb{B}^n, \mathbb{S}^{n-1})$, the retraction r cannot exist. □

Not in all spaces do partitions exist. This makes us introduce the following definition.

Definition 3. A space $\mathscr{X}$ is said to be *normal* if for any two of its disjoint closed subsets A and B there exists a set, and *completely normal* if there exists a set for any two subsets A and B having the property that the closure of each of them does not intersect with the other (such sets A and B are called *separated*).

Proposition 1. *A normal space $\mathscr{X}$ is completely normal if and only if each of its subspaces is normal.*

Proof. Clearly, every subspace of a completely normal space is completely normal (because sets separated in a subspace are separated in the entire space) and hence normal. Conversely, let A and B be separated subsets of a space $\mathscr{X}$, and let

$$O = \mathscr{X} \setminus (\bar{A} \cap \bar{B}) = (\mathscr{X} \setminus \bar{A}) \cup (\mathscr{X} \setminus \bar{B}).$$

The sets

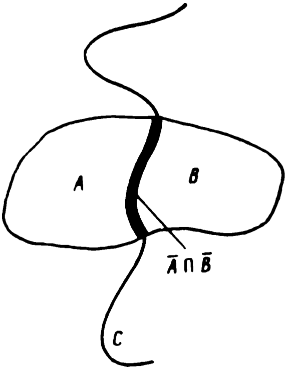
$$\bar{A} \cap O = \bar{A} \setminus (\bar{A} \cap \bar{B}) \quad \text{and} \quad \bar{B} \cap O = \bar{B} \setminus (\bar{A} \cap \bar{B})$$

are closed in O, and hence separated by some set C if O is normal. By definition $O \setminus C = U \cup V$, where U and V are disjoint sets open in O such that $\bar{A} \cap O \subset U$ and

$\bar{B} \cap O \subset V$. But since O is clearly open in $\mathscr{X}$, so are U and V, and since $A \cap \bar{B} = \varnothing$ and $B \cap \bar{A} = \varnothing$, we have $A \subset \bar{A} \cap O$ and $B \subset \bar{B} \cap O$. Hence $A \subset U$ and $B \subset V$. Since

$$O \setminus C = \mathscr{X} \setminus (C \cup (\bar{A} \cap \bar{B})),$$

this proves that the set $C \cup (\bar{A} \cap \bar{B})$ in $\mathscr{X}$ separates A and B. Thus a normal space all subspaces of which are also normal is completely normal. □



According to Proposition 1 completely normal spaces are also called *hereditary normal.*

Remark 3. By definition normal (and completely normal) spaces should have sets and be Hausdorff; but this is not the case here.

Proposition 2. *Any metric space is normal.*

Proof. Let A and B be nonintersecting closed subsets of a metric space $\mathscr{X}$. For every point $p \in A$, the number

$$\rho_B(p) = \inf_{q \in B} \rho(p, q)$$

(the *distance from p to B*) is obviously positive, and hence for that point its spherical $\rho_B(p)/2$-neighbourhood

$$U(p) = \left\{ a \in \mathscr{X},\ \rho(p, a) < \frac{\rho_B(p)}{2} \right\}$$

is defined. Similarly, for any point $q \in B$ its spherical $\rho_A(q)/2$-neighbourhood

$$V(q) = \left\{a \in \mathscr{X},\ \rho(a, q) < \frac{\rho_A(q)}{2}\right\},$$

is defined where

$$\rho_A(q) = \inf_{p \in A} \rho(p, q)$$

is the distance from q to A. Let

$$U = \bigcup_{p \in A} U(p), \quad V = \bigcup_{q \in B} V(q).$$

The sets U and V are open, contain the sets A and B, respectively, and do not intersect (if $a \in U \cap V$, then there are points $p \in A$, $q \in B$ such that $a \in U(p)$, $a \in V(q)$, and hence

$$\rho(p, a) < \frac{\rho_B(p)}{2} \leqslant \frac{\rho(p, q)}{2}$$

and

$$\rho(a, q) < \frac{\rho_A(q)}{2} \leqslant \frac{\rho(p, q)}{2},$$

but then

$$\rho(p, q) \leqslant \rho(p, a) + \rho(a, q) < \rho(p, q),$$

which is impossible). The set $C = \mathscr{X} \setminus U\ (U \cup V)$ therefore separates A and B. □

Corollary. *Any metric space is completely normal.*

Proof. It suffices to notice that every subspace of a metric space is itself a metric space (and hence normal). □

In particular, the *cube* I^n *is a completely normal space.*

Proposition 3. *Any compact Hausdorff space* $\mathscr{X}$ *is normal.*

Proof. Let A and B be closed disjoint subspaces of a space $\mathscr{X}$. Since $\mathscr{X}$ is a Hausdorff space, for any point $p \in A$ and any point $q \in B$ there exists a neighbourhood $U_q(p)$ of p and a neighbourhood $V_p(q)$ of q such that

$$U_q(p) \cap V_p(q) = \varnothing.$$

For every point $q \in B$ all its neighbourhoods $U_q(p)$, $p \in A$ form an open covering of the subspace A, and hence, since the subspace, being a closed set of a compact space, is compact, there is a finite family $p_1, \dots, p_n$ of points of the set A such that the open set

$$U_q = U_q(p_1) \cup \dots \cup U_q(p_n)$$

contains A. The set U_q does not intersect with the neighbourhood

$$V(q) = V_{p_1}(q) \cap \dots \cap V_{p_n}(q)$$

of the point q.

Neighbourhoods $V(q)$, $q \in B$, make up an open covering of a subspace B, and hence, for similar reasons, there is a finite family $q_1, \dots, q_m$ of points of the set B such that the open set

$$V = V(q_1) \cup \dots \cup V(q_m)$$

contains V. The set V does not intersect with the open set

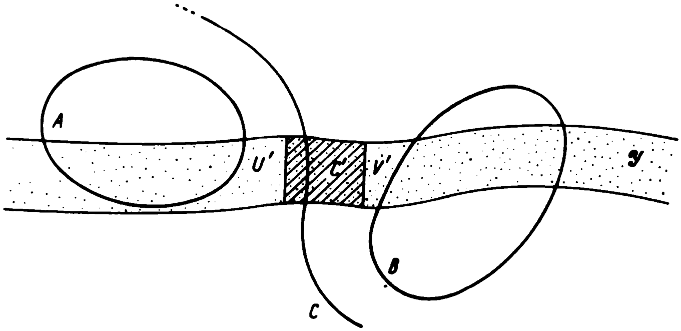
$$U = U_{q_1} \cap \dots \cap U_{q_m}$$

which contains A. The set $C = \mathscr{X} \setminus (U \cup V)$ therefore separates A and B. □

An example of a compact Hausdorff not completely normal space will be given in the next lecture.

We shall also need the following simple proposition.

Proposition 4 (separation extension). *Let $\mathscr{X}$ be a completely normal space, A and B disjoint closed subsets, and*



$\mathscr{Y}$ a closed subspace of $\mathscr{X}$. Let also C' be an arbitrary set in $\mathscr{Y}$ that separates the sets $A \cap \mathscr{Y}$ and $B \cap \mathscr{Y}$ closed

(in $\mathcal{Y}$). Then there exists a set C in $\mathcal{X}$ that separates A and B such that $C \cap \mathcal{Y} \subset C'$.

Proof. Under the hypothesis $\mathcal{Y} \setminus C' = U' \cup V'$, where U' and V' are nonintersecting sets, open (in $\mathcal{Y}$), that contain sets $A \cap \mathcal{Y}$ and $B \cap \mathcal{Y}$, respectively. Let

$$A' = A \cup U' \quad \text{and} \quad B' = B \cup V'.$$

Since U' and V' are not only open, but also closed (in $\mathcal{Y} \setminus C'$), in $\mathcal{X}$ these sets are separated. Hence the sets A' and B' are also separated, and there is therefore a set C that separates A' and B'. The same set separates A and B, of course. In addition, since

$$\mathcal{X} \setminus C \supset A' \cup B' \supset U' \cup V' = \mathcal{Y} \setminus C',$$

we have $\mathcal{Y} \cap C \subset C'$. □

Now we may return to the study of sets in the cube $\mathbf{I}^n$.

Proposition 5. *Let*

$$Q_1 \supset Q_2 \supset \ldots \supset Q_n$$

be a decreasing sequence of closed subsets of the cube $\mathbf{I}^n$ such that

Q_1 separates $\mathbf{I}_1^{n-1}(-1)$ and $\mathbf{I}_1^{n-1}(+1)$ in $\mathbf{I}^n$,
Q_2 separates $\mathbf{I}_2^{n-1}(-1) \cap Q_1$ and $\mathbf{I}_2^{n-1}(+1) \cap Q_1$ in Q_1,
. . . .
Q_n separates $I_n^{n-1}(-1) \cap Q_{n-1}$ and $\mathbf{I}_n^{n-1}(+1) \cap Q_{n-1}$ in Q_{n-1}.

Then Q_n is not empty.

Proof. Applying Proposition 4 to $\mathcal{X} = \mathbf{I}^n$, $A = \mathbf{I}_k^{n-1}(-1)$, $B = \mathbf{I}_k^{n-1}(+1)$ and $\mathcal{Y} = Q_{k-1}$, we obtain in the cube $\mathbf{I}^n$ a set C_k for any $k = 2, \ldots, n$ that separates a face $\mathbf{I}_k^{n-1}(-1)$ and a face $\mathbf{I}_k^{n-1}(+1)$ and such that

$$C_k \cap Q_{k-1} \subset Q_k.$$

For $k = 1$ we put $C_1 = Q_1$.

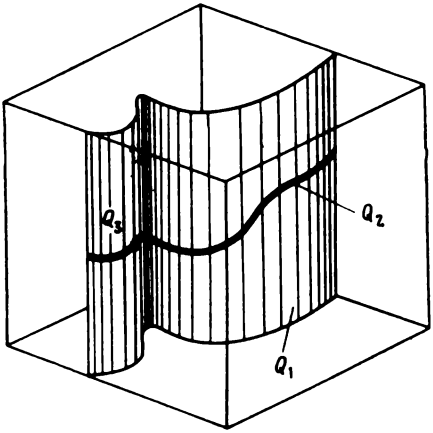
According to Theorem 3 sets $C_1, C_2, \ldots, C_n$ have the property that

$$C_1 \cap C_2 \cap \ldots \cap C_n \neq \varnothing.$$

On the other hand, since

$$C_1 = Q_1,\quad C_2 \cap Q_1 \subset Q_2,\ \ldots,\ C_n \cap Q_{n-1} \subset Q_n,$$

we have $C_1 \cap C_2 \cap \ldots \cap C_n \subset Q$. Hence $Q_n \neq \varnothing$. □



Now we can prove the main theorem of this lecture.

Theorem 4 (Lebesgue theorem on coverings of a cube). *Every finite closed covering $\{C_\alpha\}$ of I^n which consists of sets none of which intersects with any two opposite faces is of multiplicity $\geqslant n + 1$.*

Proof. Let

A_1 be the union of all elements of a covering $\{C_\alpha\}$ which intersect with the face $\mathrm{I}_1^{n-1}(-1)$ (and hence do not with the face $\mathrm{I}_1^{n-1}(+1)$),

A_2 be the union of all elements of a covering $\{C_\alpha\}$ which intersect with the face $\mathrm{I}_2^{n-1}(-1)$ and do not participate in A_1,

. . . .

A_n be the union of all elements of a covering $\{C_\alpha\}$ which intersect with the face $\mathrm{I}_n^{n-1}(-1)$ and participate neither in A_1 nor in A_2, . . ., nor in A_{n-1},

A_{n+1} be the union of all elements of a covering $\{C_\alpha\}$ which intersect with no face of the cube I^n.

Let also

$Q_1 = A_1 \cap B_1$, where $B_1 = A_2 \cup \ldots \cup A_{n+1}$,

$Q_2 = A_1 \cap A_2 \cap B_2$, where $B_2 = A_3 \cup \ldots \cup A_{n+1}$,

. . . .

$Q_{n-1} = A_1 \cap A_2 \cap \ldots \cap A_{n-1} \cap B_{n-1}$, where $B_{n-1} = A_n \cup A_{n+1}$,

$Q_n = A_1 \cap A_2 \cap \ldots \cap A_n \cap B_n$, where $B_n = A_{n+1}$.

Clearly, the sets Q_k are closed and form a decreasing sequence

$$Q_1 \supset Q_2 \supset \ldots \supset Q_n.$$

In addition, since $A_1 \cup B_1 = \mathbf{I}^n$, we have $(\mathbf{I}^n \setminus A_1) \cap (\mathbf{I}^n \setminus B_1) = \varnothing$ and since the set A_1 does not intersect with the face $\mathbf{I}_1^{n-1}(+1)$ and the sets $A_2, \ldots, A_{n+1}$ do not intersect with the face $\mathbf{I}_1^{n-1}(-1)$, we have

$$\mathbf{I}_1^{n-1}(+1) \subset \mathbf{I}^n \setminus A_1 \quad \text{and} \quad \mathbf{I}_1^{n-1}(-1) \subset \mathbf{I}^n \setminus B_1.$$

Since $\mathbf{I}^n \setminus Q_1 = (\mathbf{I}^n \setminus A_1) \cup (\mathbf{I}^n \setminus B_1)$, this proves that Q_1 separates $\mathbf{I}_1^{n-1}(-1)$ and $\mathbf{I}_1^{n-1}(+1)$ in $\mathbf{I}^n$.

Similarly, since $(A_1 \cap A_2) \cup (A_1 \cap B_2) = A_1 \cap (A_2 \cup B_2) = A_1 \cap B_1 = Q_1$, we have $(Q \setminus (A_1 \cap A_2)) \cap (Q_1 \setminus (A_1 \cap B_2)) = \varnothing$ and since the set A_2 does not intersect with the face $\mathbf{I}_2^{n-1}(+1)$ and the sets $A_3, \ldots, A_{n+1}$ do not intersect with the face $\mathbf{I}_2^{n-1}(-1)$, we have

$$\mathbf{I}_2^{n-1}(+1) \cap Q_1 \subset Q_1 \setminus (A_1 \cap A_2)$$

and

$$\mathbf{I}_2^{n-1}(-1) \cap Q_1 \subset Q_1 \setminus (A_1 \cap B_2).$$

Since $Q_2 = (A_1 \cap A_2) \cap (A_1 \cap B_2)$ and hence

$$Q_1 \setminus Q_2 = (Q_1 \setminus (A_1 \cap A_2)) \cup (Q_1 \setminus (A_1 \cap B_2)),$$

this proves that Q_2 separates $\mathbf{I}_2^{n-1}(-1) \cap Q_1$ and $\mathbf{I}_2^{n-1}(+1) \cap Q_1$ in Q_1.

Generally, if $A'_k = A_1 \cap A_2 \cap \ldots \cap A_{k+1}$ and $B'_k = A_1 \cap A_2 \cap \ldots \cap A_k \cap B_{k+1}$, where $k = 1, \ldots, n-1$, then

$$\begin{aligned} A'_k \cup B'_k &= A_1 \cap \ldots \cap A_k \cap (A_{k+1} \cup B_{k+1}) \\ &= A_1 \cap \ldots \cap A_k \cap B_k \\ &= Q_k \end{aligned}$$

and hence $(Q_k \setminus A'_k) \cap (Q_k \setminus B'_k) = \varnothing$. Also, since the set A_{k+1} and hence the set A'_k do not intersect with the face $\mathrm{I}^{n-1}_{k+1}(+1)$ and the sets $A_{k+2}, \ldots, A_{n+1}$ and hence the sets B_k and B'_k do not intersect with the face $\mathrm{I}^{n-1}_{k+1}(-1)$, we have

$$\mathrm{I}^{n-1}_{k+1}(+1) \cap Q_k \subset Q_k \setminus A'_k \text{ and } \mathrm{I}^{n-1}_{k+1}(-1) \cap Q_k \subset Q_k \setminus B'_k.$$

Since $Q_{k+1} = A'_k \cap B'_k$ and hence $Q_k \setminus Q_{k+1} = (Q_k \setminus A'_k) \cup (Q_k \setminus B'_k)$, this proves that Q_{k+1} separates $\mathrm{I}^{n-1}_{k+1}(-1) \cap Q_k$ and $\mathrm{I}^{n-1}_{k+1}(+1) \cap Q_k$ in Q_k.

Thus the sets $Q_1, \ldots, Q_n$ satisfy all conditions of Proposition 5, and hence according to this proposition the set Q_n is not empty.

Let p_0 be an arbitrary point of Q_n. Since

$$Q_n = A_1 \cap \cdots \cap A_{n+1}$$

and since each of the sets A_i is a union of elements of a covering $\{C_\alpha\}$, with elements involved in one of the sets A_k being involved in none of the others, the point p_0 is at least in $n+1$ elements of all the elements. Hence the multiplicity of the covering $\{C_\alpha\}$ is at least $n+1$. □

It is now easy to derive $\dim \mathrm{I}^n \geqslant n$ from the Lebesgue theorem.

Definition 4. An open covering $\{U_\alpha\}$ of a topological space $\mathscr{X}$ is said to be *contractible* if there is an open covering $\{V_\alpha\}$ of $\mathscr{X}$ with the same set of indices (a *contraction* of $\{U_\alpha\}$) such that

$$\bar{V}_\alpha \subset U_\alpha \quad \text{for any} \quad \alpha.$$

Proposition 6. *Each open covering of a normal space $\mathscr{X}$ is contractible.*

Proof. We prove this proposition only under the assumption that the *covering is finite*. The general case requires transfinite induction, and we do not need it here.

Let $\{U_1, \ldots, U_n\}$ be an arbitrary finite open covering of a normal space $\mathscr{X}$. Clearly, it suffices to prove that there is an open set V_1 such that $\bar{V}_1 \subset U_1$ and the family $\{V_1, U_2, \ldots, U_n\}$ is still a covering.

To this end consider the closed sets

$$\mathscr{X}\setminus U_1 \quad \text{and} \quad \mathscr{X}\setminus(U_2 \cup \dots \cup U_n).$$

Since $\{U_1, U_2, \dots, U_n\}$ is a covering, these sets do not intersect. There is therefore a set C separating the sets. Let $\mathscr{X}\setminus C = U \cup V$, where U and V are nonintersecting open sets such that

$$\mathscr{X}\setminus U_1 \subset U \quad \text{and} \quad \mathscr{X}\setminus(U_2 \cup \dots \cup U_n) \subset V.$$

The second inclusion implies that the family $\{V, U_2, \dots, U_n\}$ is a covering of the space $\mathscr{X}$, and it follows from the first inclusion that $\bar{V} \subset \mathscr{X}\setminus U \subset U_1$. It remains to put $V_1 = V$. □

Now we can prove the inequality

$$\dim \mathbf{I}^n \geqslant n.$$

Since the cube $\mathbf{I}^n$ is a compact metric space, according to Corollary 2 to Lemma 1 of the preceding lecture, this inequality implies that for some $\varepsilon > 0$ (as we shall show, any $\varepsilon < 2$ will do) every finite open ε-covering $\{U_\alpha\}$ of $\mathbf{I}^n$ is of multiplicity $\geqslant n + 1$.

Let $\{V_\alpha\}$ be an arbitrary contraction of $\{U_\alpha\}$ (existing, by Proposition 3, for finite coverings, that we have proved). Then the family $\{\bar{V}_\alpha\}$ is a closed ε-covering of the cube $\mathbf{I}^n$. Since for $\varepsilon < 2$ this covering clearly satisfies the condition of the Lebesgue theorem, by this theorem the covering $\{\bar{V}_\alpha\}$ is of multiplicity not less than $n + 1$. But then the original covering $\{U_\alpha\}$ is also of multiplicity not less than $n + 1$. □

Lecture 10

Ordinals. Interval topology in sets of ordinals· Zero-dimensional spaces· Tychonoff's example· Tychonoff product of topological spaces· Filters. Centred sets of sets· Ultrafilters· Compactness criterions· The Tychonoff theorem

---

This lecture dealing entirely with general topology, which has nothing to do with the theory of smooth manifolds, consists of two parts interrelated by the name of A. Tychonoff. In the first part we construct Tychonoff's example of a Hausdorff normal, but not completely normal compact zero-dimensional space that contains a one-dimensional subspace. The main goal of the second part is to prove the Tychonoff theorem on products of compact spaces.

Tychonoff's example is constructed from the so-called transfinite ordinals; we briefly recall their definition and basic properties.

A set A is called a *partially ordered set* if a relation $\leqslant$ is defined on it, which has the property of being *reflexive* ($a \leqslant a$ for any $a \in A$), *antisymmetric* (if $a \leqslant b$ and $b \leqslant a$, then $a = b$) and *transitive* (if $a \leqslant b$ and $b \leqslant c$, then $a \leqslant c$). If $a \leqslant b$ but $a \neq b$, then we write $a < b$. A mapping $\varphi: A \to B$ of partially ordered sets is said to be *monotone* if $\varphi a \leqslant \varphi b$ when $a \leqslant b$. A bijective monotone mapping whose inverse is also monotone is called an *isomorphism*. Partially ordered sets A and B are said to be *isomorphic* if there exists at least one isomorphism $A \to B$. The isomorphism of partially ordered sets is an equivalence relation, and therefore all partially ordered sets break down into classes of isomorphic sets.

A partially ordered set A is said to be *ordered* (or, more extensively, *linearly ordered*) if for any two elements $a, b \in A$ either $a \leqslant b$ or $b \leqslant a$. Classes of isomorphic ordered sets are termed *ordinal* *types*. A mapping of ordered sets is an isomorphism if and only if it is bijective and monotone.

An ordered set A is said to be *well ordered* if each of its nonempty subsets (including the set A itself) has a first element. Ordinal types of well ordered sets are termed *ordinal numbers* (or *ordinals*). A well ordered set A belonging to an ordinal number α is said to *have type* α or to be a *set of type* α. It is also said that α is an *ordinal number of the set* A.

If ordered sets A and B are of the same type α, then they have equal power. Their cardinality is the *cardinality of* α and is denoted by $|\alpha|$.

If a set A is finite, then its ordinal number is called *finite* (or an ordinal number of the *first number class*). Since for any natural number n any two ordered sets of cardinality n are isomorphic, from the equation $|\alpha| = |\beta|$ it follows for finite ordinal numbers α and β that $\alpha = \beta$. Such numbers can therefore be identified with their cardinalities, i.e. with the natural numbers

$$0,\ 1,\ 2,\ \ldots,\ n,\ \ldots \tag{1}$$

(Natural numbers are equally suitable both for count and recount.)

If a set A is countable, then its ordinal number α is called *countable* (or an ordinal number of the *second number class*). An example of a countable ordinal number is the ordinal type of the set (1) of all natural numbers (obviously of a well ordered set). This number is denoted by ω.

For any element a_0 of a well ordered set A, the set of all elements $a \in A$ for which $a < a_0$ is termed a *segment* (*initial*) of a set A, which is defined by the element a_0 and denoted by $A(a_0)$. It is also well ordered.

If every set of type β is isomorphic to some section of a set of type α, then we say that β is *smaller* than α and write $\beta < \alpha$. By definition the formula $\beta \leqslant \alpha$ means that either $\beta < \alpha$ or $\beta = \alpha$.

Clearly, the *relation* $<$ *is transitive* (if $\alpha < \beta$ and $\beta < \gamma$, then $\alpha < \gamma$).

The set of all ordinal numbers β for which $\beta < \alpha$ is denoted by $[0, \alpha]$ and termed an *interval with end point* α. Similarly the set of all ordinals β with $\beta \leqslant \alpha$ is denoted by $[0, \alpha]$ and termed a *segment with end point* α. Clearly, $[0, \alpha] = [0, \alpha) \cup \{\alpha\}$.

For example, an interval $[0, \omega)$ is the set (1) of all natural numbers, and a segment $[0, \omega]$ is the set (1), with the element ω added on the right.

Since any section of the set (1) is of the form $[0, n]$, we see, in particular, that ω is a *smallest ordinal number of the second number class*.

It is easy to see that *no well ordered set* A *can be isomorphic to a section* $B(a_0)$ *of any of its subsets* $B \subset A$ (*the case* $B = A$ *is not excluded*). Indeed, reasoning by reductio ad absurdum we assume that the isomorphism $\varphi\colon A \to B(a_0)$ exists and consider the set of all elements $a \in A$ for which $\varphi a < a$. The set contains a_0, and is therefore not empty. Let a_1 be its first element (it exists because A is a well ordered set), and let $a_2 = \varphi a_1$. Then $a_2 < a_1$, and simultaneously $\varphi a_2 < \varphi a_1 = a_2$, which contradicts the choice of the element a_1. This contradiction proves that the isomorphism φ cannot exist. □

For $B = A$ we have, in particular, that *if* $\beta < \alpha$, *then* $\beta \neq \alpha$.

It is also easy to see that *every interval* $[0, \alpha)$ *is a well ordered set of type* α. Indeed, associating each element a of a well ordered set A of type α with the type of the section $A\ (a)$ we obviously obtain an isomorphic mapping $A \to [0, \alpha)$. □

It follows that there is a *first element in every set A of ordinal numbers*. Indeed, choose in A an arbitrary element α_0. If α_0 is a first element, then there is nothing to prove, and if α_0 is not a first element, then the intersection $[0, \alpha_0) \cap A$ is not empty, and the first element of the intersection (which exists because the interval $[0, \alpha_0)$ is well ordered) is a first element of the set A as well. □

Since $[0, \alpha)$ is a well ordered interval, so is the segment $[0, \alpha] = [0, \alpha) \cup \{\alpha\}$. The ordinal type of this segment is designated $\alpha + 1$. Since the interval $[0, \alpha)$ is a section $[0, \alpha]\ (\alpha)$ of the segment $[0, \alpha]$, we have $\alpha < \alpha + 1$. In addition, *if* $\beta \leqslant \alpha + 1$, *then* $\beta \leqslant \alpha$ (for any section of the segment $[0, \alpha]$ is contained in the interval $[0, \alpha)$). For this reason the number $\alpha + 1$ is said to *succeed* the number α.

An ordinal β is said to be *limiting* if there is no ordinal number α such that $\beta = \alpha + 1$. An example of a limiting ordinal is the ordinal ω.

For any ordinals α and β, consider the intersection $D = [0, \alpha) \cap [0, \beta)$ of the intervals $[0, \alpha)$ and $[0, \beta)$. It turns out that *there exists an ordinal number* $\delta \leqslant \alpha$ *such that* $D = [0, \delta)$. Indeed, if $D = [0, \alpha)$, then $\delta = \alpha$. Let $D \neq [0, \alpha)$, i.e. the complement $C = [0, \alpha) \setminus D$ of the set D in $[0, \alpha)$ is not empty, and let δ be the first element of the complement. If $\xi \in [0, \delta)$, then $\xi < \alpha$ (since $\xi < \delta$ and $\delta < \alpha$) and $\xi \notin C$ (since $\xi < \delta$ and δ is a first element in C), i.e. $\xi \in D$. Conversely, if $\xi \in D$ and $\delta \leqslant \xi$, then $\delta < \beta$ (since $\xi < \beta$), and hence $\delta \in D$, which is impossible. Hence $\xi < \delta$, i.e. $\xi \in [0, \delta)$. □

By symmetry, along with $\delta \leqslant \alpha$ the inequality $\delta \leqslant \beta$ also holds. If $\delta < \alpha$ and $\delta < \beta$, then $\delta \in D = [0, \delta)$, which is impossible. Therefore either $\delta = \alpha$ (and then $\alpha \leqslant \beta$) or $\delta = \beta$ (and then $\beta \leqslant \alpha$). This proves that for any two ordinal numbers α and β either $\alpha = \beta$ or $\alpha < \beta$ or $\beta < \alpha$, i.e. *every set of ordinal numbers is ordered and hence* (since each of its subsets has a first element) *well ordered.*

Hence it is easy to see that *for any subset B of an arbitrary ordered set A the type* β *of the set B does not exceed the type* α *of the set A*:

$$\beta \leqslant \alpha.$$

Indeed, otherwise there would hold an inequality $\alpha < \beta$, and the set A would be isomorphic to some section of the subset B, which as we know is impossible. □

Suppose we are given an arbitrary family $\{\alpha_\xi\}$ of ordinal numbers indexed by numbers ξ of some interval $[0, \tau)$, and let A_ξ be an arbitrary set of type α_ξ. We order a disjoint union

$$A = \bigsqcup_{\xi} A_\xi$$

of all sets A_ξ by assuming that for elements $a \in A_\xi$ and $b \in A_\eta$ the inequality $a < b$ holds if and only if either $\xi < \eta$ or $\xi = \eta$ and $a < b$ in the set A_ξ. For any nonempty subset $C \subset A$ the set of all numbers $\xi < \tau$ for which $C \cap A_\xi \neq \varnothing$ is a nonempty subset of the interval $[0, \tau)$, and hence has a first element ξ_0. Clearly, the first element of the set $C \cap A_{\xi_0}$ is also the first element of the entire set C. This proves that A is a *well ordered set* and hence its type α is an ordinal number. Obviously the number α depends only on the numbers α_ξ (and is independent of the choice of the set A_ξ). It is called the *sum* of the numbers α_ξ and designated

$$\sum_{\xi<\tau} \alpha_\xi .$$

Notice that α *depends* on the order of indexing the numbers α_ξ. (For example, since the set of type $\omega + n$ has a last element and that of type $n + \omega$ has not, we have $n + \omega \neq \omega + n$.)

Since $A_\xi \subset A$ for any ξ, we have $\alpha_\xi \leqslant \alpha$. (The equality is quite possible here; for example, it is clear that $n + \omega = \omega$.)

However, $\alpha_\xi < \alpha + 1$. Since any set of ordinal numbers, being well ordered, has some type τ and hence its elements can be indexed by numbers of the interval $[0, \tau)$, this proves that *for any set of ordinal numbers, there exists a number greater than all of its elements.*

Hence it follows that the *notion of a set of all ordinal numbers is meaningless.* An attentive reader will have noticed that in formulating our statements somewhat akwardly we have avoided mentioning that particular "set". Now he knows the reason why.

In particular, there exist ordinal numbers that are greater than all ordinal numbers of the second (or *third number class*). The least of them is the *first ordinal number of the third number class*, and is denoted by Ω.

For any countable (or finite) sequence $\{\alpha_n\}$ of the second number class their sum $\sum\limits_{n<\omega} \alpha_n$ is obviously an ordinal number of the second number class (for a disjoint union of countable sets is countable). Therefore the least of the ordinal numbers greater than all of the ordinal numbers α_n (this ordinal number is designated $\sup \alpha_n$) is also an ordinal number of the second number class (in shorter form, *if* $\alpha_n < \Omega$, *then* $\sup \alpha_n < \Omega$).

The existence of the number Ω proves that there exist uncountable ordinal numbers. In fact *there exist ordinal numbers of any cardinality*, i.e., which is obviously the same, *any set A can be indexed by ordinal numbers of some interval* $[0, \alpha)$. To prove this statement (known as the Zermelo theorem) we say that β is a *distinguished* ordinal number if there is a subset of the set A that can be indexed by numbers of an interval $[0, \beta)$. The set B of all distinguished ordinal numbers is not empty (for example, if A is an infinite set, then all ordinal numbers of the first and second number class are distinguished) and has the property that if $\beta \in B$ and $\gamma < \beta$, then $\gamma \in B$, i.e. $[0, \beta) \subset B$.

Let $*$ be an element which is not in A, and let $A^* = A \sqcup \{*\}$. Construct a mapping $\varphi: B \to A^*$ according to the following rules:

(a) By choosing an arbitrary element $a_0 \in A$, put $\varphi(0) = a_0$.

(b) Suppose that for some ordinal number $\beta \in B$ the mapping φ has already been constructed on an interval $[0, \beta)$, and let $A' = \varphi[0, \beta)$. If $A' \subset A$ and $A' \neq A$, then we arbitrarily choose an element $a_\beta \in A \setminus A'$ and put

$$\varphi(\beta) = a_\beta.$$

Otherwise (i.e. when either $A' = A$ or $* \in A'$) we put $\varphi(\beta) = *$.

Clearly, this construction defines φ on the whole B. [The set B' of all $\beta \in B$ for which $\varphi(\beta)$ is not defined, being a set of ordinal numbers, has, if not empty, a first element $\beta_0 \in B'$. Then the mapping φ is defined on $[0, \beta_0)$, and hence by (b) the element $\varphi(\beta)$ is also defined, i.e. $\beta_0 \notin B'$. The obtained contradiction proves that $B' = \varnothing$.]

Let B^* be a set, possibly empty, of all ordinal numbers β for which $\varphi(\beta) = *$, and let $\alpha = \sup(B \setminus B^*)$. Then by construction φ maps the interval $[0, \alpha)$ onto the set A, and this mapping is bijective, i.e. φ specifies the indexing of the set A by the numbers of the interval $[0, \alpha)$. □

Every set A of ordinal numbers is a topological space with respect to the so-called *interval topology* whose base for open sets are the intersections of the set A with intervals of the form (α, β), where $\alpha < \beta$ (by definition every interval (α, β) consists of all numbers γ for which $\alpha < \gamma < \beta$).

This space is obviously Hausdorff.

Remark 1. The interval topology can be introduced on any ordered set; for example, on a set $[0, \Omega) \times [0, 1)$ ordered *lexicographically* (i.e. so that $(\alpha_1, t_1) < (\alpha_2, t_2)$ if and only if $\alpha_1 < \alpha_2$ or $\alpha_1 = \alpha_2$ and $t_1 < t_2$). The resulting topological space is called an *Alexandroff half-line*. It can be thought of as a result of glueing in a closed interval $[0, 1]$ between any two adjacent ordinal numbers α and $\alpha + 1$ of the second number class.

Remark 2. It can be shown that a long half-line, with the initial point $(0, 0)$ removed, has a natural C^ω-smoothness with respect to which it is a one-dimensional Hausdorff manifold which does not satisfy the second countability axiom. (Cf. Remark 5 of Lecture 7.)

The interval topology on sets of ordinal numbers has a remarkable and somewhat unexpected property that *for any ordinal number* ξ *the segment* $[0, \xi]$ *is a compact space in the interval topology.* Indeed, let $\mathfrak{U}$ be an arbitrary open covering of $[0, \xi]$. We call a number $\eta \leqslant \xi$ *admissible* if the segment $[0, \eta]$ can be covered by a finite subfamily of $\mathfrak{U}$. The set of all admissible numbers is not empty (it contains 0), and if it does not exhaust the entire segment $[0, \xi]$, then due to ordinal numbers being well ordered, there exists a least inadmissible number $\gamma \leqslant \xi$. Let U_0 be an element of $\mathfrak{U}$ containing γ and let (α, β) be an interval such that $\gamma \in (\alpha, \beta) \subset U_0$. Since $\alpha < \gamma$, α is an admissible number, and hence the segment $[0, \alpha]$ can be covered by a finite subfamily of $\mathfrak{U}$. By adding an element U_0 to that subfamily we obtain a finite subfamily which covers the interval $[0, \beta) = [0, \alpha) \cup (\alpha, \beta)$ and hence the segment $[0, \gamma] \subset [0, \beta)$. Consequently, in spite of the assumption γ is an admissible number. The obtained contradiction shows that all numbers of the segment $[0, \xi]$ are admissible. In particular, so is the number ξ. Consequently, the covering $\mathfrak{U}$ contains a finite subcovering. □

The topological space $\mathscr{X} = [0, \xi]$ has the property that *each of its points has a base for the neighbourhood system that is not only an open but also a closed set* (we say about such spaces $\mathscr{X}$ that ind $\mathscr{X} = 0$). To prove this it suffices to notice that if β is not a limiting ordinal number (i.e. $\beta = \gamma + 1$), then $(\alpha, \beta) = [\alpha + 1, \gamma]$ for any α. □

Moreover, a similar statement is true not only for points but also for any closed sets, i.e. *for every closed set* $F \subset [0, \xi]$ *and any of its neighbourhoods* U *there exists a neighbourhood* $V \subset U$ *which is a closed set* (about such spaces $\mathscr{X}$ we say that Ind $\mathscr{X} = 0$). Indeed, according to the above said every point $\alpha \in F$ has a closed neighbourhood V_α contained in U. Neighbourhoods V_α make up an open covering of the set F from which, since F, being a closed subset of a compact space, is compact, we can choose a finite subcovering $\{V_{\alpha_1}, \ldots, V_{\alpha_n}\}$. The union $V =$

$V_{\alpha_1} \cup \ldots \cup V_{\alpha_n}$ of elements of that subcovering is precisely the closed neighbourhood of a subset F contained in U. $\square$

Thus we have proved for an arbitrary topological space $\mathscr{X}$ that *if* $\mathscr{X}$ *is a compact space and* ind $\mathscr{X} = 0$, *then* Ind $\mathscr{X} = 0$. [If $\mathscr{X}$ *is therewith a Hausdorff space and* Ind $\mathscr{X} = 0$, *then* ind $\mathscr{X} = 0$. To prove this it suffices to notice that in a Hausdorff space any one-point set is closed.]

On the other hand, it is easy to see that *if* Ind $\mathscr{X} = 0$, *then* dim $\mathscr{X} = 0$. Indeed, the equation dim $\mathscr{X} = 0$ implies that any finite open covering $\{U_1, \ldots, U_n\}$ consisting of nonintersecting open sets and hence (since the complement of any element V_j of the covering $\{V_1, \ldots, V_m\}$ is the union of all the other elements of the covering and is therefore open) having the property that any of its elements V_j is not only open but also closed, may be a refinement of every finite open covering $\{U_1, \ldots, U_m\}$.

We prove (for a space $\mathscr{X}$ with Ind $\mathscr{X} = 0$) even more, namely, that the *number m of elements of the covering* $\{V_1, \ldots, V_m\}$ *may be considered equal to the number n of elements of the covering* $\{U_1, \ldots, U_n\}$, *and that this covering can be constructed so that for any* $i = 1, \ldots, n$ *there holds an inclusion* $V_i \subset U_i$.

For $n = 1$, put $V_1 = U_1$.

Suppose a covering $\{V_1, \ldots, V_n\}$ is constructed for any covering $\{U_1, \ldots, U_n\}$ consisting of $n \geqslant 1$ open sets. Consider an arbitrary open covering $\{U_1, \ldots, U_n, U_{n+1}\}$ consisting of $n + 1$ sets. Under the induction hypothesis applied to the covering $\{U_1, \ldots, U_{n-1}, U_n \cup U_{n+1}\}$ there exists an open covering $\{V_1, \ldots, V_{n-1}, V'_n\}$ consisting of n nonintersecting sets such that

$$V_1 \subset U_1, \ldots, V_{n-1} \subset U_{n-1}, \quad V'_n \subset U_n \cup U_{n+1}.$$

On the other hand, since the set

$$F = \mathscr{X} \setminus (U_1 \cup \ldots \cup U_n)$$

is closed and contained in U_{n+1}, by virtue of the condition Ind $\mathscr{X} = 0$ there exists an open and at the same time closed set V such that $F \subset V \subset U_{n+1}$. We put

$$V_n = V'_n \cap (\mathscr{X} \setminus V), \quad V_{n+1} = V'_n \cap V.$$

To complete the proof it remains to notice that all sets $V_1, \ldots, V_{n-1}, V_n, V_{n+1}$ are open, do not intersect, cover $\mathscr{X}$ and have the property that $V_i \subset U_i$ for any $i = 1, \ldots, n+1$. □

In particular, we see that *for any ordinal number* ξ

$$\dim [0, \xi] = 0.$$

It turns out that not only from $\operatorname{Ind} \mathscr{X} = 0$ it follows that $\dim \mathscr{X} = 0$, but conversely from $\dim \mathscr{X} = 0$ it follows that $\operatorname{Ind} \mathscr{X} = 0$ (so that $\dim \mathscr{X} = 0$ *if and only if* $\operatorname{Ind} \mathscr{X} = 0$). Indeed, if $\dim \mathscr{X} = 0$ and if U is a neighbourhood of a closed set $F \subset \mathscr{X}$, then since the sets U and $\mathscr{X} \setminus F$ make up a finite open covering of a space $\mathscr{X}$, a finite covering $\{V_j\}$ consisting of nonintersecting open, and hence closed sets, can be a refinement of that covering. Then the union V of all elements of the covering V_j which intersect with F (and hence lie in U) is an open and at the same time closed neighbourhood of F contained in U. □ T

It is now easy to prove that *if* $\mathscr{X}$ *and* $\mathscr{Y}$ *are Hausdorff and compact spaces and* $\dim \mathscr{X} = 0$ *and* $\dim \mathscr{Y} = 0$, *then*

$$\dim (\mathscr{X} \times \mathscr{Y}) = 0$$

(so that for these spaces formula (8) of Lecture 8 is true). Indeed, by virtue of the above said on the relation between ind, Ind, and dim for spaces $\mathscr{X}$ and $\mathscr{Y}$ equations $\operatorname{ind} \mathscr{X} = 0$ and $\operatorname{ind} \mathscr{Y} = 0$ hold, and for the space $\mathscr{X} \times \mathscr{Y}$ the equality $\dim (\mathscr{X} \times \mathscr{Y}) = 0$ is equivalent to the equality $\operatorname{ind} (\mathscr{X} \times \mathscr{Y}) = 0$. On the other hand, it is clear that *if* $\operatorname{ind} \mathscr{X} = 0$ *and* $\operatorname{ind} \mathscr{Y} = 0$, *then* $\operatorname{ind} (\mathscr{X} \times \mathscr{Y}) = 0$ (for if U is a closed neighbourhood of a point p in $\mathscr{X}$ and V is a closed neighbourhood of a point q in $\mathscr{Y}$, then the product $U \times V$, having an open complement $(\mathscr{X} \setminus U)\, \mathscr{Y} \cup \mathscr{X} \times (\mathscr{Y} \setminus V)$, is a closed neighbourhood of a point (p, q) in $(\mathscr{X} \times \mathscr{Y})$. □

Remark 3. The notation suggests that along with the invariant dim (and the "improved" invariant dim′) the other integer invariants ind and Ind can be defined which can also pretend to playing the part of dimension. This is really so, but unfortunately these invariants coincide only for metrizable spaces satisfying the second countability axiom and diverge outside them.

The study of these and other similar invariants is the subject of the so-called **dimension theory.**

After the preliminary work we may set directly to constructing the Tychonoff example.

Let $\mathscr{X} = [0, \omega] \times [0, \Omega]$ be a direct product of segments of ordinal numbers $[0, \omega]$ and $[0, \Omega]$. According to the above results $\mathscr{X}$ is a zero-dimensional ($\dim \mathscr{X} = 0$) Hausdorff and compact space. Let A be an open subspace of $\mathscr{X}$, which consists of all points except for the "angular point" (ω, Ω). We show that $\dim A \geqslant 1$. Since the inequality $\dim A \geqslant 1$ implies that $\dim A \neq 0$, and because $\dim A = 0$ is equivalent to $\operatorname{Ind} A = 0$, it suffices to prove that $\operatorname{Ind} A \neq 0$, i.e. that there is in A a closed set F and a neighbourhood U of F such that no neighbourhood V of F contained in U is a closed set. We prove even more, namely, we find in A a closed set F and a neighbourhood U of F such that *for any neighbourhood V of the set A contained in U its closure $\overline{V}$ is not contained in U.* (This means that there is no set between F and $A \setminus U$ in A, i.e. that A is not normal. On the other hand, being compact and Hausdorff, the space $\mathscr{X}$ is normal. *The space $\mathscr{X}$ is* thus *an example of a normal but not of a completely normal space.*)

We take as F the set of all points of $\mathscr{X}$ of the form (n, Ω), where $0 \leqslant n < \omega$, and as U the set of all points of the form (n, β), where $0 \leqslant n < \omega$ and $0 \leqslant \beta \leqslant \Omega$. Clearly, the set U is open (in A and in $\mathscr{X}$), and the set F is closed (in A but not in $\mathscr{X}$). Let V be an arbitrary open set (in A and hence in $\mathscr{X}$) which contains F and is contained in U. For any n, $0 \leqslant n < \omega$, the set of all numbers $\beta \leqslant \Omega$ for which $(n, \beta) \in V$ is open in $[0, \Omega]$ (why?) and contains Ω. There is a number $\alpha_n < \Omega$ therefore such that the entire half-interval $(\alpha_n, \Omega] = (\alpha_n, \Omega) \cup \{\Omega\}$ is contained in that set. Let $\alpha = \sup \alpha_n$. Since $\alpha_n < \Omega$, we have $\alpha < \Omega$ (see above), and for any number $\beta \in (\alpha, \Omega)$ every point (n, β) is in V.

On the other hand, every neighbourhood W of a point (ω, β) in A contains a neighbourhood of the form $[N, \omega] \times$

(β_1, β_2), where N is some natural number $(<\omega)$ and β_1 and β_2 are ordinal numbers of the second number class such that $\beta_1 < \beta < \beta_2$. The neighbourhood W therefore contains all the points (n, β) with $n > N$ and hence intersects with V. Consequently, $(n, \omega) \in \bar{V}$ and hence $\bar{V} \not\subset U$. □

Let us now return to direct products and carry over their construction to the case of any (in general, infinite) number of factors.

Let $\{\mathscr{X}_\alpha, \alpha \in A\}$ be an arbitrary family of topological spaces, and let $\mathscr{X}$ be the set of all possible families $\{x_\alpha\}$, where $x_\alpha \in \mathscr{X}_\alpha$.

For any **finite** set of indices $\{\alpha_1, \ldots, \alpha_n\}$ and any open subsets $U_{\alpha_1} \in \mathscr{X}_{\alpha_1}, \ldots, U_{\alpha_n} \subset \mathscr{X}_{\alpha_n}$ we denote by $O(U_{\alpha_1}, \ldots, U_{\alpha_n})$ a subset of $\mathscr{X}$ consisting of all points $\{x_\alpha\}$ for which

$$x_{\alpha_1} \in U_{\alpha_1}, \ldots, x_{\alpha_n} \in U_{\alpha_n}.$$

Clearly

$$O(U_{\alpha_1}, \ldots, U_{\alpha_n}) = O(U_{\alpha_1}) \cap \ldots \cap O(U_{\alpha_n}),$$

and hence the intersection of any two sets of the form $O(U_{\alpha_1}, \ldots, U_{\alpha_n})$ is also a set of the same form. Therefore the set of all sets $O(U_{\alpha_1}, \ldots, U_{\alpha_n})$ is a base for open sets of some topology on a set $\mathscr{X}$ (open sets of the topology are thus all possible unions of sets of the form $O(U_{\alpha_1}, \ldots, U_{\alpha_n})$).

Definition 1. The resulting topological space $\mathscr{X}$ is called the *direct* (or *Tychonoff*) *product* of spaces $\mathscr{X}_\alpha$. It is designated as

$$\prod_{\alpha \in A} \mathscr{X}_\alpha$$

(also denoted by $\bigtimes_{\alpha \in A} \mathscr{X}_\alpha$). If $\{\mathscr{X}_\alpha\}$ is a finite family, we obviously obtain a direct product in the sense of Definition 5 of Lecture 8.

Theorem 1. *The direct product of any family of compact spaces is compact.*

This theorem is known as the **Tychonoff theorem**. To prove it, we give another definition of a compact space in terms convenient for proving the theorem.

Let for the time being $\mathscr{X}$ be an arbitrary set.

Definition 2. A set Φ of subsets of $\mathscr{X}$ is said to be a *filter* if

1° any subset of $\mathscr{X}$ containing some subset of Φ is in Φ:

$$A \subset B \subset \mathscr{X}, \quad A \in \Phi \Rightarrow B \in \Phi,$$

2° the intersection of any two subsets of Φ is in Φ:

$$A, B \in \Phi \Rightarrow A \cap B \in \Phi,$$

3° all $\mathscr{X}$ are in Φ, but an empty set $\varnothing$ is not:

$$\varnothing \notin \Phi, \quad \mathscr{X} \in \Phi.$$

Example 1. For any point x_0 of a set $\mathscr{X}$ a family of all subsets of $\mathscr{X}$ containing the point x_0 is a filter.

Example 2. If $\mathscr{X}$ is a topological space, then for any point $x_0 \in \mathscr{X}$ the family of all subsets A of $\mathscr{X}$ for which x_0 is an interior point ($x_0 \in \operatorname{Int} A$) is a filter.

Example 3. If $\mathscr{X}$ is an infinite set, then all sets whose complements are finite form a filter.

Example 4. If Φ is a filter and M is a subset which is not in Φ, then the set Φ_M of all subsets $A \subset \mathscr{X}$ for which $A \cup M \in \Phi$ forms a filter.

Definition 3. A set Γ of subsets of a set $\mathscr{X}$ is said to be *centred* if the intersection of any of its finite subfamilies is not empty.

Clearly, *if* Γ *is centred, then for any mapping* $f\colon \mathscr{X} \to \mathscr{Y}$ *the set* $f\Gamma = \{fA;\ A \in \Gamma\}$ *is also centred.*

From Properties 2 and 3 it follows immediately that *any filter is a centred set.*

Conversely, it is easy to see that *any centred set* Γ *is contained in some filter*. Indeed, it is clear that such a filter is, for example, the set $\Phi(\Gamma)$ of all subsets of $\mathscr{X}$ that have the property that each of them contains a subset which is the intersection of some finite family of elements of Γ. $\square$

The filter Φ (Γ) is obviously a minimal filter containing the centred set Γ. It is said to be *generated* by this set.

Recall that an element m of a partially ordered set A is said to be *maximal* if it follows from $m \leqslant a$ that $m = a$. Elements a and b of A are said to be *comparable* if either $a \leqslant b$ or $b \leqslant a$. A subset C of a partially ordered set is said to be a *chain* if it is linearly ordered, i.e. any two of its elements are comparable. An element c_0 is called an *upper bound* of the chain C if $c \leqslant c_0$ for any element $c \in C$. A partially ordered set A is called an *inductively* ordered set if any of its chains has at least one upper bound.

An example of a partially ordered set is the set FILTR $(\mathscr{X})$ of all filters on $\mathscr{X}$ which is ordered by the relation of inclusion ($\Phi \leqslant \Psi$ if $\Phi \subset \Psi$).

Maximal elements of the set are called *ultrafilters* or *maximal filters*. A filter Φ is thus an ultrafilter if any filter that contains it coincides with Φ.

It is easy to see that a *set Φ of subsets of $\mathscr{X}$ is an ultrafilter if and only if*:

(a) *the intersection of any two sets of Φ is not empty*,

(b) *the entire set $\mathscr{X}$ is in Φ*,

(c) *of any two mutually complementary subsets A and $B = \mathscr{X} \setminus A$ of $\mathscr{X}$ one (and only one) is in Φ.*

Indeed, condition (a) is Condition 2° in the definition of a filter, and condition (b) is the first part of Condition 3° the second part of which follows from the first and condition (c) (applied to $A = \mathscr{X}$). Further, if $A \subset B \subset \mathscr{X}$ and $A \in \Phi$ but $B \notin \Phi$, then by virtue of condition (c) $\mathscr{X} \setminus B \in \Phi$, which contradicts condition (a) since $(\mathscr{X} \setminus B) \cap A = \varnothing$. Hence the set of subsets Φ that satisfies conditions (a) to (c) is a filter. If the filter is contained in a filter Ψ and there is a set $A \in \Psi$ which is not in Φ, then $B = \mathscr{X} \setminus A$ is in Φ and hence in Ψ, which is impossible (since $A \cap B = \varnothing$). Therefore $\Psi = \Phi$ and hence Φ is an ultrafilter.

Conversely, if Φ is an ultrafilter, then conditions (a) and (b) are obviously satisfied, and in order to prove condition (c) it suffices to notice that for $A \notin \Phi$ the filter Φ_A (see Example 4) obviously contains the filter Φ.

Therefore, since the ultrafilter Φ is maximal, the filter Φ_A coincides with Φ. But it is clear that the filter Φ_A contains the set $B = \mathscr{X} \setminus A$. Hence $B \in \Phi$. $\square$

It follows immediately that *every filter in Example 1 is an ultrafilter*. Such ultrafilters ase called *trivial.*

Further, it is easy to see that *any set A that intersects with each element of an ultrafilter Φ is in Φ*. Indeed, if $A \notin \Phi$, then in the ultrafilter Φ there is an element (namely $\mathscr{X} \setminus A$) that does not intersect with A. $\square$

Chains of the set FILTR ($\mathscr{X}$) are the filter sets $\{\Phi_\alpha\}$ such that for any α and β either $\Phi_\alpha \subset \Phi_\beta$ or $\Phi_\beta \subset \Phi_\alpha$. It is easy to see that the *union Φ of all filters of a chain is a filter* (it suffices to notice that any two elements of the union are in some filter Φ_α). Since the filter Φ contains all the filters Φ_α it is their upper bound. Thus we see that the *set of all filters is inductive.*

But according to the well-known **Zorn lemma** (also known as **Kuratowski-Zorn theorem**) *any inductively ordered set A has a maximal element and, what is more, for any element $a_0 \in A$ there is a maximal element m such that $a_0 \leqslant m$.*

For completeness we give a proof of the lemma. Let a set A be arbitrarily indexed by ordinal numbers of some interval $[0, \alpha)$. (Note that this indexing is in no way connected with partial ordering in A.) It is clear that we may assume without loss of generality that a given element a_0 has a number 0. Consider in A a subset C satisfying the following conditions:

(a) the element a_0 is in C,

(b) an element of A having a number ξ is in C if and only if it is comparable with all elements of A having smaller numbers.

It is clear that these conditions characterize uniquely a subset C and that subset is a chain. Let m be an arbitrary upper bound of the set C (which exists due to the induction condition). Since the element m is comparable with any element of C, every element $a \geqslant m$ is also comparable with any element of C and, in particular, with all elements of C that have smaller numbers. By virtue of condition (b) therefore the element a is in C and hence $a \leqslant m$. This proves that m is a maximal element. To complete the proof it remains to notice that by construction $a_0 \leqslant m$.

By virtue of the Zorn lemma applied to an inductively ordered set FILTR ($\mathscr{X}$) *any filter on a set $\mathscr{X}$ is contained in some ultrafilter.*

This gives us an unlimited supply of nontrivial ultrafilters.

Remark 4. It is curious that no nontrivial ultrafilter can be constructed explicitly without any arbitrary elements, i.e., in other words, characterized by a list of properties so that it is the only ultrafilter with these properties. (In mathematical logic the proof of this statement is given on the basis of explicating clearly all the necessary notions of mathematical logic.)

Now let $\mathscr{X}$ be a topological space.

Definition 4. A point $p_0 \in \mathscr{X}$ is said to be *an adherent point* of a set Φ of subsets of a topological space $\mathscr{X}$ if every neighbourhood of that point intersects with every element A of Φ, i.e. in other words, if $p_0 \in \bar{A}$ for any $A \in \Phi$.

Thus adherent points of the set Φ are the points of intersection

$$\bigcap_{A \in \Phi} \bar{A} \tag{2}$$

of closures $\bar{A}$ of all sets $A \in \Phi$.

Notice that *if Φ is a filter, then set* (2) *coincides with the intersection of all closed sets of* Φ.

The complement of set (2) is precisely the union of open sets $\mathscr{X} \setminus \bar{A}$. Consequently, a *set Φ has no adherent points if and only if the sets $\mathscr{X} \setminus \bar{A}$, $A \in \Phi$, cover $\mathscr{X}$.*

Definition 5. If every neighbourhood of a point $p_0 \in \mathscr{X}$ contains some element of a set of subsets of Φ (and hence, when Φ is a filter, is in Φ), Φ is said to *converge to the point p_0*.

It is easy to see that if a *set Φ is an ultrafilter, then it converges to each of its adherent points.* Indeed, if a neighbourhood U of a point p_0 intersects with each element of the ultrafilter Φ, then U is in Φ. □

Now we can prove the compactness criterion we need.

Proposition 1. *The following properties of a topological space $\mathscr{X}$ are equivalent:*

1° *The space $\mathscr{X}$ is compact.*

2° *Any centred set of its closed subsets has a nonempty intersection.*

3° *Every filter on $\mathscr{X}$ has an adherent point.*

4° Every ultrafilter on $\mathscr{X}$ converges.

Proof. It obviously suffices to prove the implications

$$\begin{tikzcd} & & 3^\circ \arrow[dd, Rightarrow] \\ 1^\circ \arrow[r, Leftrightarrow] & 2^\circ \arrow[ur, Rightarrow] & \\ & & 4^\circ \arrow[ul, Rightarrow] \end{tikzcd}$$

Implications 1° ⇒ 2° and 2° ⇒ 1°. The statement that the intersection of all elements A of some set Γ of subsets of a space $\mathscr{X}$ is empty implies that the complements $\mathscr{X} \setminus A$ of these elements cover $\mathscr{X}$ and the statement that Γ is a centred set implies that no finite family of complements $\mathscr{X} \setminus A$ covers $\mathscr{X}$. Therefore if a set of closed subsets of a compact space $\mathscr{X}$ has an empty intersection, then it cannot possibly be centred, and if the intersection of any centred set of closed subsets of the space $\mathscr{X}$ is not empty, then every family of open sets of $\mathscr{X}$, no finite subfamily of which covers $\mathscr{X}$, does not cover $\mathscr{X}$.

Implication 2° ⇒ 3°. It suffices to notice that the set Γ of all closed subsets which are in an arbitrary filter Φ is centred and that every adherent point of Φ is a common point of elements of Γ.

Implication 3° ⇒ 4°. If every filter has adherent points, then, in particular, every utralfilter also has an adherent point. On the other hand, an ultrafilter converges if and only if it has adherent points.

Implication 4° ⇒ 2°. Every centred set is contained in some ultrafilter. Therefore, if the ultrafilter converges, then the centred set has an adherent point (any adherent point of an arbitrary set of sets is an adherent point of every subset of that set as well). On the other hand, if a set of sets consists of closed sets, then each of its adherent points is in the intersection of all its elements. Thus, if in a space $\mathscr{X}$ every ultrafilter converges (has an adherent point), then any centred set of closed subsets of $\mathscr{X}$ has a nonempty intersection. □

Now the Tychonoff theorem can be easily proved.

Proof of Theorem 1. Let Φ be an arbitrary ultrafilter on a direct product

$$\mathscr{X} = \prod_{\alpha \in A} \mathscr{X}_\alpha$$

of compact spaces $\mathscr{X}_\alpha$. Consider the image $\mathrm{pr}_\alpha \Phi$ of the ultrafilter under the canonical projection $\mathrm{pr}_\alpha\colon \mathscr{X} \to \mathscr{X}_\alpha$, $\{x_\alpha\} \mapsto x_\alpha$. Since Φ is a centred set, its projection $\mathrm{pr}_\alpha \Phi$ is also centred. So is therefore a set Γ_α which consists of closures of elements of the set $\mathrm{pr}_\alpha \Phi$ (i.e. of sets $\bar{A}_\alpha$, where $A_\alpha = \mathrm{pr}_\alpha A$, $A \in \Phi$). Consequently, since a space $\mathscr{X}_\alpha$ is compact, there exists a point $p_\alpha \in \mathscr{X}_\alpha$ which belongs to all sets $\bar{A}_\alpha$ of Γ_α, i.e. such that every its neighbourhood U_α intersects with every set A_α.

For the sets A of the filter Φ this implies that they all intersect with the set $O(U_\alpha)$, from which it follows, since Φ is an ultrafilter, that $O(U_\alpha) \in \Phi$.

Choosing for α one of the points p_α and thus obtaining some point $\{p_\alpha\}$ of the product $\mathscr{X}$ consider an arbitrary neighbourhood W of that point in $\mathscr{X}$. By definition the neighbourhood contains some set of the form $O(U_{\alpha}, \ldots, U_{\alpha_n})$, where $U_{\alpha_1}, \ldots, U_{\alpha_n}$ are neighbourhoods of the points $p_{\alpha_1}, \ldots, p_{\alpha_n}$ in the spaces $\mathscr{X}_{\alpha_1}, \ldots, \mathscr{X}_{\alpha_n}$ respectively. But, as we have just seen, all sets $O(U_{\alpha_1}), \ldots, O(U_{\alpha_n})$ are in the ultrafilter Φ. So is their intersection

$$O(U_{\alpha_1}, \ldots, U_{\alpha_n}) = O(U_{\alpha_1}) \cap \ldots \cap O(U_{\alpha_n}),$$

and hence the neighbourhood W.

Thus any neighbourhood of the point $\{p_\alpha\}$ is in the ultrafilter Φ, and hence Φ converges to $\{p_\alpha\}$.

This proves that any ultrafilter converges in the space $\mathscr{X}$. Hence $\mathscr{X}$ is a compact space. $\square$

Lecture 11

Smoothness on an affine space· The manifold of matrices of a given rank· Stiefel manifolds· Matrix rows· The exponential of a matrix· The logarithm of a matrix· Orthogonal and J-orthogonal matrices· Matrix Lie groups· Groups of J-orthogonal matrices· Unitary and J-unitary matrices· Complex matrix Lie groups· Complex-analytic manifolds· Arcwise connected spaces· Connected spaces· The coincidence of connectedness and arcwise connectedness for manifolds· Smooth and piecewise smooth paths· Connected manifolds failing to satisfy the second axiom of countability

---

Returning to smooth C^∞-manifolds, we supplement the list of examples given in Lecture 6. These examples not only illustrate the general theory but are also a source of important theorems which unfortunately are almost entirely beyond the scope of this course.

Let $\mathscr{A}$ be an arbitrary n-dimensional affine space over the field $\mathbb{R}$. Every frame $O\mathbf{e}_1 \ldots \mathbf{e}_n$ of the space defines a coordinate isomorphism $h: \mathscr{A} \to \mathbb{R}^n$ (carrying a point M with $\mathbf{OM} = x^1\mathbf{e}_1 + \ldots + x^n\mathbf{e}_n$ into a point $(x^1, \ldots, x^n)$ of the space $\mathbb{R}^n$) and hence a chart $(\mathscr{A}, h)$.

This chart makes up a one-element atlas on $\mathscr{A}$ and, consequently, gives on $\mathscr{A}$ a smoothness (cf. Example 1 of Lecture 6). For any other frame $O'\mathbf{e}'_1 \ldots \mathbf{e}'_n$ the corresponding chart $(\mathscr{A}, h')$ is associated with $(\mathscr{A}, h)$ via a mapping of transition $h' \circ h^{-1}$ which is an affine mapping $\mathbb{R}^n \to \mathbb{R}^n$ (and hence is a diffeomorphism) of class C^ω, and is therefore compatible with $(\mathscr{A}, h)$. This means that the smoothness given by the chart $(\mathcal{A}, h)$ is independent of the choice of the frame $O\mathbf{e}_1 \ldots \mathbf{e}_n$. We call it the *standard smoothness* on $\mathcal{A}$. It is obviously compatible with the topology on $\mathcal{A}$.

Since every vector space is trivially an affine space, all this applies, in particular, to an arbitrary vector space; for example, to the mn-dimensional vector space $\mathbb{R}(m, n)$ of all real rectangular m by n matrices. Thus $\mathbb{R}(m, n)$ *is an mn-dimensional smooth C^ω-manifold.* To obtain a chart to cover this manifold it is necessary to write out the elements of each matrix A of $\mathbb{R}(m, n)$ in a row in an arbitrary order the same for all matrices.

A group GL (n) of all nonsingular n by n matrices is distinguished in the vector space $\mathrm{Mat}_n(\mathbb{R}) = \mathbb{R}(n, n)$ of all square matrices A of order n by the condition $\det A \neq 0$. Since the determinant of a matrix, being a polynomial in its elements, is continuously dependent on them (the function $\det\colon A \to \det A$ is continuous on $\mathrm{Mat}_n(\mathbb{R})$), the group GL (n) is an open set of the smooth manifold $\mathrm{Mat}_n(\mathbb{R})$, and hence an induced smoothness is defined on it. Thus the *group* GL (n) *is a smooth C^ω-manifold.* The dimension of this manifold is n^2.

To obtain a less trivial example of a smooth manifold we introduce into consideration a subset $\mathbb{R}(m, n; k)$ of the space $\mathbb{R}(m, n)$ which consists of matrices of rank k, where $0 \leqslant k \leqslant \min(m, n)$. Being a subset of the topological space $\mathbb{R}(m, n)$ this set is a topological space (with respect to the induced topology).

It is found possible to *introduce in an ordinary way into a set $\mathbb{R}(m, n; k)$ a smoothness compatible with the topology, relative to which it is a smooth manifold of dimension $k(m + n - k)$.*

To this end, consider a subset $U \subset \mathbb{R}(m, n; k)$ consisting of all m by n matrices of the form

$$(1)\qquad \begin{Vmatrix} A & B \\ C & D \end{Vmatrix},$$

where A is a nonsingular k by k matrix (i.e. a matrix from GL (k)) and B, C, and D are arbitrary matrices of order

$(m-k)$ by k, k by $(n-k)$, or $(m-k)$ by $(n-k)$, respectively.

By choosing and fixing a certain order of writing out elements of the matrices A, B, and C in a single row we obtain a mapping h of U onto some subset U_0 of a space $\mathbb{R}^N$, where

$$N = k^2 + (m-k)\,k + k\,(n-k) = k\,(m+n-k).$$

Since matrix (1) has the same rank as the matrix

$$\begin{Vmatrix} E & 0 \\ -CA^{-1} & E \end{Vmatrix} \begin{Vmatrix} A & B \\ C & D \end{Vmatrix} = \begin{Vmatrix} A & B \\ 0 & D-CA^{-1}B \end{Vmatrix}$$

(for the first factor on the left is nonsingular), and the last matrix has the rank k if and only if $D - CA^{-1}B = 0$, matrix (1) is uniquely defined by matrices A, B, C, where A is a nonsingular matrix and no conditions are in general imposed on B and C. This means that the set U and hence U_0 are open (in $\mathbb{R}(m, n; k)$ and $\mathbb{R}^N$, respectively), and the mapping $h: U \to U_0$ is bijective. Consequently, the pair (U, h) is a chart in $\mathbb{R}(m, n; k)$.

The inverse mapping $h^{-1}: U_0 \to U$ consists in reconstructing the matrices A, B, C by a given row vector and filling in the bottom right corner of matrix (1) with matrix

$$D = CA^{-1}B.$$

Now let $\alpha: \mathbb{R}(m, n; k) \to \mathbb{R}(m, n; k)$ be a bijective mapping of the set $\mathbb{R}(m, n; k)$ onto itself given by some interchange of rows or columns (there are $m!n!$ such mappings all together), and let

$$U_\alpha = \alpha U \quad \text{and} \quad h_\alpha = h \circ \alpha^{-1}.$$

Then the pair (U_α, h_α) is also a chart in $\mathbb{R}(m, n; k)$.

Since any m by n matrix of rank k can be carried into a matrix of form (1) by interchanging its rows and columns the *sets* U_α *cover the set* $\mathbb{R}(m, n; k)$.

Let $U_\alpha \cap U_\beta \neq \varnothing$. The set

$$\alpha^{-1}(U_\alpha \cap U_\beta) = U \cap (\alpha^{-1} \circ \beta)\, U$$

consists of matrices (1) which, after an interchange of $\alpha^{-1} \circ \beta$ are still in U, i.e. whose minor of order k changes under $\alpha^{-1} \circ \beta$ into the upper left-hand minor is nonzero. This minor is some rational (since $D = CA^{-1}B$) and hence continuous (in its domain of definition) function of elements of the matrices A, B, C. In $U_0 = h(U)$ therefore the set on which this function is nonzero, i.e. the set

$$h(\alpha^{-1}(U_\alpha \cap U_\beta)) = h_\alpha(U_\alpha \cap U_\beta),$$

is open.

The mapping

$$(2) \quad h_\beta \circ h_\alpha^{-1} = h \circ (\beta^{-1} \circ \alpha) \circ h^{-1}\colon h_\alpha(U_\alpha \cap U_\beta) \to h_\beta(U_\alpha \cap U_\beta)$$

carries a point of $h_\alpha(U_\alpha \cap U_\beta) \subset U_0$, corresponding to matrix (1) (more precisely, to the matrices A, B, and C) first to matrix (1), i.e. to the matrix

$$\begin{Vmatrix} A & B \\ C & CA^{-1}B \end{Vmatrix},$$

then subjects this matrix to some interchange of rows and columns, and finally writes in a single row all elements found in the place of the matrices A, B, and C. This shows that mapping (2) is given by rational functions of the coordinates. These functions are defined on $h_\alpha(U_\alpha \cap U_\beta)$ and hence are smooth on $h_\alpha(U_\alpha \cap U_\beta)$. This proves that *any two charts of the form* (U_α, h_α) *are compatible with each other.*

Consequently, all these charts make up some atlas. The smoothness defined by this atlas (which is obviously compatible with the topology in $\mathbb{R}(m, n; k)$) is precisely the smoothness we wished to construct. □

Notice that when $m = n = k$ a single chart makes up an atlas, and it is precisely the atlas which specified the above smoothness on GL (n).

Let $\mathscr{V}$ be an arbitrary n-dimensional vector space over the field of real numbers. Assuming as *m-frames* in $\mathscr{V}$, where $0 \leqslant m \leqslant n$, linearly independent families of vectors from $\mathscr{V}$ consisting of m vectors, consider the set $V(m, \mathscr{V})$ of all m-frames in $\mathscr{V}$. Choosing in $\mathscr{V}$ an arbitrary basis (an n-frame), we can associate each m-frame with an m by n matrix whose columns consist of the coordinates of the vectors of that frame in the given basis. It is clear that this gives some bijective mapping $V(m, \mathcal{V}) \to \mathbb{R}(m, n; m)$. Carrying with the aid of this mapping the smoothness from $\mathbb{R}(m, n; m)$ to $V(m, \mathcal{V})$ we obtain on $V(m, \mathcal{V})$ the structure of a smooth manifold which is obviously independent of the choice of basis in $\mathcal{V}$.

This manifold is called the *Stiefel manifold* of the vector space $\mathcal{V}$. When $\mathcal{V} = \mathbb{R}^n$ it is denoted by $V(m, n)$.

Problem 1. Show that the set $G(m, \mathcal{V})$ of all m-dimensional subspaces of an n-dimensional vector space $\mathcal{V}$ is an $m(n-m)$-dimensional smooth manifold. (This manifold is called a *Grassmann manifold*; cf. Lecture II.10.)

To study more complex smooth manifolds consisting of matrices we need some facts from the general matrix theory.

On a vector space $\mathrm{Mat}_n(\mathbb{R})$ of all square n by n matrices $A = \| a_i^j \|$ one can introduce many different norms, i.e. functions $A \mapsto |A|$ with values in the field $\mathbb{R}$ such that

$$|A+B| \leqslant |A| + |B|, \quad |kA| = |k| \cdot |A|$$

and

$$|A| > 0 \text{ for } A \neq 0$$

for any matrices A, B and any number k. The norm defined by the formula

$$|A| = n \cdot \max_{1 \leqslant i,\, j \leqslant n} |a_i^j|$$

will suit. It has the property that *for any matrices* $A = \| a_j^i \|$ *and* $B = \| b_j^i \|$ *we have an inequality*

$$|AB| \leqslant |A| \cdot |B|.$$

Indeed, by the definition of matrix multiplication the elements c_j^i of the matrix AB can be expressed as $c_j^i = a_k^i b_j^k$, i.e. as

$$c_j^i = \sum_{k=1}^{n} a_k^i b_j^k, \qquad i, j = 1, \ldots, n.$$

Hence

$$|AB| = n \cdot \max_{1 \leqslant i,\, j \leqslant n} |c_j^i| \leqslant n \cdot \max_{1 \leqslant i,\, j \leqslant n} \sum_{k=1}^{n} |a_k^i| \cdot |b_j^k|$$

$$\leqslant n \cdot \underbrace{\left(\frac{|A|}{n} \cdot \frac{|B|}{n} + \cdots + \frac{|A|}{n} \cdot \frac{|B|}{n} \right)}_{n \text{ times}} = |A| \cdot |B|. \square$$

Moreover, for this norm

$$|A^{\top}| = |A|.$$

In accordance with general calculus definitions we shall say that a sequence $\{A(m)\}$ of matrices $A(m) = \| a^i_j(m) \|$ *converges* to a matrix $A = \| a^i_j \|$ (and write $A(m) \to A$ as $m \to \infty$) if

$$|A(m) - A| \to 0$$

as $m \to \infty$ or, what is clearly equivalent, if $a^i_j(m) \to a^i_j$ for any $i, j = 1, \ldots, m$.

An infinite matrix series

$$A_0 + A_1 + \ldots + A_m + \ldots \tag{3}$$

is said to *converge* to a matrix A if a sequence $\{A(m)\}$ of its *partial sums*

$$A(m) = A_0 + \ldots + A_m$$

converges to A.

The well-known Cauchy's test for convergence can be referred directly to matrix series, i.e. *series* (3) *converges if and only if for any* $\varepsilon > 0$ *there is* $M_0 > 0$ *such that for any* $m > M_0$ *and every* $p \geqslant 0$

$$|A_m + A_{m+1} + \ldots + A_{m+p}| < \varepsilon.$$

Series (3) is said to be *absolutely convergent* if a number series

$$|A_0| + |A_1| + \ldots + |A_m| + \ldots$$

is convergent.

Since

$$|A_m + A_{m+1} + \ldots + A_{m+p}| \leqslant |A_m| + |A_{m+1}| + \ldots + |A_{m+p}|$$

it follows directly from Cauchy's test for convergence that *every absolutely convergent series converges.*

An example of an absolutely convergent series is a series

$$\sum_{m=0}^{\infty} \frac{A^m}{m!} = E + A + \frac{A^2}{2!} + \ldots + \frac{A^m}{m!} + \ldots$$

This series is (absolutely) convergent for any matrix A. Its sum is called the *exponential* function of a matrix and designated exp A or e^A.

Clearly the operation of transition to the matrix e^A commutes with the operation $A \mapsto A^{\top}$ of transposition of matrices, i.e.

$$(e^A)^{\top} = e^{A^{\top}} \text{ for any matrix } A. \tag{4}$$

In addition,

$$C^{-1}e^AC = e^{C^{-1}AC} \tag{5}$$

for any nonsingular matrix C (since $C^{-1}A^mC = (C^{-1}AC)^m$ for every m).

If a matrix A is an upper triangular matrix with diagonal elements $a_1, \ldots, a_n$, then every matrix A^m is also upper triangular, and its diagonal elements are equal to the mth powers $a_1^m, \ldots a_n^m$ of diagonal elements of the matrix A. Hence the matrix e^A is also upper triangular, and its diagonal elements are the numbers $e^{a_1}, \ldots, e^{a_n}$. Hence follows the formula for the determinant of e^A

$$\det e^A = e^{\operatorname{Tr} A}, \tag{6}$$

where $\operatorname{Tr} A$ is the trace of the matrix A (the sum of its diagonal elements). But we know from linear algebra (since the trace is one of the coefficients of a characteristic polynomial) that for every nonsingular matrix

$$\operatorname{Tr}(C^{-1}AC) = \operatorname{Tr} A \tag{7}$$

(see Proposition 5 of Lecture II.5) and any matrix is of the form $C^{-1}AC$, where A is some upper triangular matrix (generally, with complex elements). By virtue of (5) therefore *formula* (6) *is valid for any matrix* A.

In particular, we see that *every matrix of the form* e^A *has a positive determinant*, so that, in particular, the *matrix* e^A *is nonsingular* (is in the group GL (n)).

For arbitrary matrices A and B the equation

$$e^Ae^B = e^{A+B} \tag{8}$$

is in general false. *It is true*, however, *if matrices* A *and* B *commute*, because in that case the calculation proving formula (8) for numbers remains valid. [By the binomial formula for any $m \geqslant 0$

$$\frac{(A+B)^m}{m!} = \sum_{k=0}^{m} \frac{A^k}{k!}\,\frac{B^{m-k}}{(m-k)!},$$

and therefore for any $M > 0$

$$\sum_{m=0}^{2M} \frac{(A+B)^m}{m!} = \left(\sum_{m=0}^{M} \frac{A^m}{m!}\right)\left(\sum_{m=0}^{M} \frac{B^m}{m!}\right) + R_M,$$

where R_M is a sum of terms of the form

$$\frac{A^k}{k!}\cdot\frac{B^l}{l!}$$

taken over all pairs (k, l) for which $k + l \leqslant 2M$ and one of the numbers is greater than M. But since the number of such pairs is equal to $M(M+1)$ and

$$\left|\frac{A^k}{k!}\cdot\frac{B^l}{l!}\right| \leqslant \frac{|A|^k}{k!}\,\frac{|B|^l}{l!} \leqslant \frac{\max(|A|,\ |B|)^{2M}}{M!},$$

we have

$$|R_M| \leqslant M(M+1)\,\frac{\max(|A|,\ |B|)^{2M}}{M!}.$$

Hence $|R_M| \to 0$ as $M \to \infty$, which proves formula (8).]

In particular, we see that $e^A e^{-A} = e^0 = E$, i.e.

$$(9) \qquad (e^A)^{-1} = e^{-A}.$$

(Notice that this again proves the inclusion $e^A \in \mathrm{GL}(n)$.)

Similar to exponential functions of matrices we can define other functions of matrix A. Consider, for example, a series

$$(10) \qquad (A-E) - \frac{(A-E)^2}{2} + \ldots + (-1)^{m+1}\frac{(A-E)^m}{m} + \ldots,$$

which is the matrix analogue of a familiar **logarithmic series.** Since a logarithmic series is absolutely convergent provided $|z-1|<1$ *series* (10) *is absolutely convergent provided* $|A-E|<1$. Its sum is designated $\ln A$ and called the *logarithm of matrix* A.

Note that the *logarithm* $\ln A$ *is thus defined only when* $|A-E|<1$.

Generally, $\ln(AB) \neq \ln A + \ln B$, but *if matrices* A *and* B *commute* (and matrices $\ln A$, $\ln B$, and $\ln(AB)$ are defined), *then*

$$(11) \qquad \ln(AB) = \ln A + \ln B.$$

The proof is similar to the proof of formula (6).

If A is a number (a 1 by 1 matrix), then it is known that $e^{\ln A} = A$. Substituting the series $\ln A$ in the series e^A for A, we therefore, after removing the brackets and collecting like terms, obtain A (give a direct proof of the statement), these calculations being valid for any A for which the series $\ln A$ converges. Since these calculations (together with justification of its validity) remain valid for matrices of any order, we see therefore that

$$(12) \qquad e^{\ln A} = A$$

for any matrix A *with* $|A-E|<1$. (In particular, for any such matrix $\det A > 0$.)

For similar reasons we must also have

$$(13) \qquad \ln e^A = A,$$

the *equation* (13) *holding for* $|A| < \ln 2$ since for numbers the corresponding calculations with series are valid only for $|z| <$

$\ln 2$ (why?). (Notice that if $|A| < \ln 2$, then $|e^A - E| \leqslant e^{|A|} - 1 < 1$.)

It follows from (12) and (13) that the *exponential functions and logarithms of a matrix give mutually inverse bijections between the neighbourhood of a unit matrix E consisting of matrices A of the form e^B, where $|B| < \ln 2$, and the neighbourhood of a zero matrix 0 consisting of matrices B for which $|B| < \ln 2$.*

Since the elements of matrices e^A and $\ln A$ are obviously smooth (real analytic) functions of the elements of matrix A, *these mappings are diffeomorphisms.*

Problem 2. Derive from (8) formula (11). [**Hint.** Prove that for $|A| < \ln 2$ and $|B| < \ln 2$ the commutativity of matrices e^A and e^B yields the commutativity of matrices A and B.]

Does formula (8) follow from (11)?

Remark 1. While, unless otherwise stated, we are mainly concerned with *real matrices* (consisting of real numbers), we shall sometimes have to consider matrices with complex elements as well (for example, see above the proof of formula (6)). It is worth noting therefore that what we have said earlier of matrices e^A and $\ln A$ is equally true both for real and complex matrices.

Recall (see Lecture I.14) that a real matrix A is said to be *orthogonal* if a transposed matrix $A^{\top}$ is the inverse of it, i.e. if $A^{\top}A = E$.

Orthogonal n by n matrices form a group which is designated $\mathrm{O}(n)$.

Generally, if J is an arbitrary nonsingular n by n matrix, then the n by n matrix A for which

$$A^{\top}JA = J \tag{14}$$

is said to be a *J-orthogonal matrix*. (Thus E-orthogonal matrices are precisely orthogonal matrices.) J-orthogonal matrices form a group which is designated $\mathrm{O}_J(n)$.

It follows from (14) that the determinant of a J-orthogonal matrix A is equal to ± 1. A J-orthogonal matrix A with $\det A = 1$ is called *unimodular* or *proper*.

Unimodular J-orthogonal matrices form a subgroup $\mathrm{SO}_J(n)$ of the group $\mathrm{O}_J(n)$ (a subgroup $\mathrm{SO}(n)$ for $J = E$).

If for a J-orthogonal matrix A an inequality $|A - E| < 1$ holds and a matrix $B = \ln A$ therefore exists then

$$e^{JBJ^{-1}} = Je^{B}J^{-1} = JAJ^{-1} = (A^{\top})^{-1} = (A^{-1})^{\top}$$
$$= (e^{-B})^{\top} = e^{-B^{\top}}.$$

Consequently, if $|B| < \ln 2$ (and hence $|JBJ^{-1}| < \ln 2$ and $|-B^{\mathsf{T}}| = |B| < \ln 2$), then $JBJ^{-1} = -B^{\mathsf{T}}$, i.e.

$$B^{\mathsf{T}}J = -JB. \tag{15}$$

Conversely, if (15) holds, then

$$(e^B)^{\mathsf{T}} = e^{B^{\mathsf{T}}} = e^{-JBJ^{-1}} = Je^{-B}J^{-1} = J(e^B)^{-1}J^{-1},$$

and hence the matrix $A = e^B$ is a J-orthogonal matrix.

Matrices satisfying condition (15) are called *J-skew-symmetric*. (For $J = E$ they are the usual skew-symmetric matrices.)

Since relation (15) is linear in B, *J-skew-symmetric matrices (of a given order n) make up a linear subspace of the space* $\mathrm{Mat}_n(\mathbb{R})$. This subspace is designated by a symbol $\mathfrak{so}_J(n)$ (and $\mathfrak{so}(n)$ for $J = E$).

Consider a general situation.

Let $\mathcal{G}$ be a group consisting of n by n real matrices (a subgroup of the group $\mathrm{GL}(n) = \mathrm{GL}(n, \mathbb{R})$).

Definition 1. A group $\mathcal{G}$ is said to be a *matrix Lie group* if there is a linear subspace $\mathfrak{g}$ in the vector space $\mathrm{Mat}_n(\mathbb{R})$ such that

1° For any matrix $B \in \mathfrak{g}$ the matrix e^B is in the group $\mathcal{G}$.

2° If $e^B \in \mathcal{G}$ and $|B| < \ln 2$, then $B \in \mathfrak{g}$.

The vector space $\mathfrak{g}$ is called the *Lie algebra* of the group $\mathcal{G}$. (This name is explained by the fact that this group is an algebra in the general algebraic sense relative to an operation $[B_1, B_2] = B_1B_2 - B_2B_1$, see Remark 1 of Lecture 16 below.)

Thus, according to the above computations, *for any nonsingular matrix J the group $\mathrm{O}_J(n)$ is a matrix Lie group with Lie algebra $\mathfrak{so}_J(n)$.*

So is the entire group $\mathrm{GL}(n, \mathbb{R})$ (for this group $\mathfrak{g} = \mathrm{Mat}_n(\mathbb{R})$, notice that for this reason the vector space $\mathrm{Mat}_n \mathbb{R}$ is also designated $\mathfrak{gl}(n)$). The subgroup $\mathrm{GL}^+(n)$ of $\mathrm{GL}(n)$ consisting of matrices with positive determinant (for this subgroup too $\mathfrak{g} = \mathrm{Mat}_n(\mathbb{R})$) is also a matrix Lie group.

It follows immediately from (6) that for the group $\mathrm{SL}(n) = \mathrm{SL}(n; \mathbb{R})$ of all unimodular matrices Properties 1° and 2° are satisfied by the $(n^2 - 1)$-dimensional subspace $\mathfrak{sl}(n)$ of all matrices B for which $\mathrm{Tr}\, B = 0$ (physicists call them matrices *without spurs*). The *group* $\mathrm{SL}(n)$ *is thus a matrix Lie group with Lie algebra* $\mathfrak{sl}(n)$.

The significance of matrix Lie groups for the theory of smooth manifolds is determined by the following proposition.

Proposition 1. *Every matrix Lie group $\mathcal{G}$ has a natural structure of a smooth manifold.*

The dimension of this manifold is equal to the dimension of the algebra $\mathfrak{g}$.

Proof. Let U_0 be the neighbourhood of a zero matrix 0 in a vector space $\mathfrak{g}$ consisting of all matrices $B \in \mathfrak{g}$ for which $|B| < \ln 2$, and let $U = \exp U_0$. Then U is open in $\mathcal{G}$ (why?) and the mapping $\ln\colon A \to \ln A$ is a bijective (and even homeomorphic) mapping $U \to U_0$.

Choosing a certain way of writing elements of matrices of $\mathfrak{g}$ in a row and thus identifying the algebra $\mathfrak{g}$ with the space $\mathbb{R}^N$, where $N = \dim \mathfrak{g}$, we may think of U_0 as a subset of $\mathbb{R}^N$. Clearly, this subset is open in $\mathbb{R}^N$, and hence the pair $(U, \ln)$ is a chart in $\mathfrak{g}$.

Notice that $E \in U$.

For any matrix C of $\mathcal{G}$ we introduce into consideration a set $CU \subset \mathcal{G}$ consisting of all matrices of $\mathcal{G}$ of the form CA, where $A \in U$, and the mapping

$$h_C\colon CU \to U_0$$

of that set onto a set $U_0 \subset \mathbb{R}^N$ defined as follows:

$$h_C(CA) = \ln A, \quad A \in U.$$

Clearly, the mapping h_C is bijective, i.e. the pair (CU, h_C) is a chart in $\mathcal{G}$. Since $C = CE \in CU$, all charts of the form (CU, h_C) cover $\mathcal{G}$. On the other hand, for any two matrices $C_1, C_2 \in \mathfrak{g}$ the mapping

$$h_{C_2} \circ h_{C_1}^{-1}\colon h_{C_1}(C_1U \cap C_2U) \to h_{C_2}(C_1U \cap C_2U)$$

is given as follows:

$$(h_{C_2} \circ h_{C_1}^{-1})(B) = \ln (C_2^{-1}C_1e^B),$$

with the set $h_{C_1}(C_1U \cap C_2U)$ consisting of all matrices $B \in U_2$ for which $C_2^{-1}C_1e^B \in U$.

Therefore the set $h_{C_1}(C_1U \cap C_2U)$ is open, and the mapping $h_{C_2} \circ h_{C_1}^{-1}$ is smooth. Hence the charts (C_1U, h_{C_1}) and (C_2U, h_{C_2}) are compatible.

We thus see that charts of the form (CU, h_C) make up an atlas on $\mathcal{G}$ and hence give on $\mathcal{G}$ the structure of a smooth manifold (obviously of dimension N).

This completes the proof of Proposition 1. □

In what follows we always assume that every matrix Lie group $\mathcal{G}$ is provided with the constructed smoothness.

Being a subset of the vector space $\mathrm{Mat}_n(\mathbb{R})$, the group $\mathcal{G}$ inherits from $\mathrm{Mat}_n(\mathbb{R})$ its topology.

It is clear that in this topology all charts (CU, h_C) satisfy the condition of Proposition 2 of Lecture 7, and hence the *smoothness in a group $\mathcal{G}$ is compatible with the topology of the group.*

We thus see, in particular, that the *group* SL (n) *is a smooth manifold of dimension* $n^2 - 1$.

Similarly, *for any nonsingular matrix J of order n the group* $\mathrm{O}_J(n)$ *of all J-orthogonal n by n matrices is a smooth manifold.*

The dimension of that manifold (equal to that of the Lie algebra $\mathfrak{so}_J(n)$) is in general dependent on J.

For example, for $J = E$ the algebra $\mathfrak{so}_E(n) = \mathfrak{so}(n)$ consists of all skew-symmetric matrices $B = \|b_j^i\|$ of order n. Since in the n^2-dimensional vector space $\mathrm{Mat}_n(\mathbb{R})$ of all n by n matrices skew-symmetric matrices satisfy $n(n+1)/2$ independent conditions

$$b_j^i = -b_j^i, \text{ where } i, j = 1, \ldots, n \quad \text{and} \quad i \leqslant j,$$

we have

$$\dim \mathfrak{so}(n) = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}.$$

The *dimension of the Lie group* O (n) *is* thus *equal to* $\frac{n(n-1)}{2}$.

Another important example results when the number n is even ($n = 2m$) and the matrix J is of the form

$$J = \begin{Vmatrix} 0 & E \\ -E & 0 \end{Vmatrix}, \tag{16}$$

where 0 and E are zero and unit matrices of order m. In this case the condition that the n by n matrix

$$A = \begin{Vmatrix} A_1 & A_2 \\ A_3 & A_4 \end{Vmatrix} \tag{17}$$

(where A_1, A_2, A_3, A_4 are m by m matrices) should be J-skew-symmetric is equivalent to three relations

$$A_1^\top = -A_4, \quad A_2^\top = A_2, \quad A_3^\top = A_3 \tag{18}$$

implying that matrices A_2 and A_3 are symmetric and that matrix A_4 can be expressed in terms of A_1 (subject to no conditions). The dimension of the space of J-skew-symmetric matrices (which in this case is designated $\mathfrak{sp}(m, \mathbb{R})$) is thus equal to

$$m^2 + 2 \cdot \frac{m(m+1)}{2} = m(2m+1) = \frac{n(n+1)}{2}.$$

Matrices J-orthogonal with respect to matrix (16) are called *symplectic matrices*, and their group $O_J(n)$ (usually designated $\mathrm{Sp}(m, \mathbb{R})$) is called a *symplectic group*. According to the computation the *group* $\mathrm{Sp}(m, \mathbb{R})$ *is a smooth manifold of dimension* $m(2m+1)$.

Matrices of even order $n = 2m$, which are simultaneously orthogonal and symplectic, make up an *orthogonal symplectic group* $\mathrm{Sp}(m, \mathbb{R}) \cap O(2m)$. This is a matrix Lie group with Lie algebra $\mathfrak{sp}(m, \mathbb{R}) \cap \mathfrak{so}(2m)$.

Clearly, matrix (17) is skew-symmetric if and only if $A_1^\top = -A_1$, $A_2^\top = -A_3$, and $A_4^\top = -A_4$. Together with conditions (18) this states that *matrix* (17) *belongs to the Lie algebra* $\mathfrak{sp}(m, \mathbb{R}) \cap \mathfrak{so}(2m)$ *if and only if*

$$A_1^\top = -A_1, \quad A_2^\top = A_2 \quad \text{and} \quad A_3 = -A_2, \quad A_4 = A_1.$$

Therefore

$$\dim(\mathfrak{sp}(m, \mathbb{R}) \cap \mathfrak{so}(2m)) = \frac{m(m-1)}{2} + \frac{m(m+1)}{2} = m^2,$$

and hence the *group* Sp $(m, \mathbb{R}) \cap$ O $(2m)$ *is a smooth manifold of dimension* m^2.

Let $n = p + q$, and let

$$J = \begin{Vmatrix} E_p & 0 \\ 0 & -E_q \end{Vmatrix},$$

where E_p and E_q are unit matrices of order p and q, respectively.

Matrices J-orthogonal with respect to this matrix J are called *pseudo-orthogonal matrices of type* (p, q), and their group $\mathrm{O}_J(n)$ is designated $\mathrm{O}(p, q)$. (Thus $\mathrm{O}(n, 0) = \mathrm{O}(0, n) = \mathrm{O}(n)$.) Clearly, the *group* $\mathrm{O}(p, q)$ *is a smooth manifold of dimension* $\frac{n(n-1)}{2}$ (independent of p and q).

J-unitary matrices with complex elements satisfying the relation

$$\bar{A}^{\mathsf{T}} J A = J \tag{19}$$

are the complex analogue of J-orthogonal matrices (these are usual unitary matrices when $J = E$, see Lecture II.21), and the *J-skew-Hermitian matrices* which satisfy the relation

$$\bar{B}^{\mathsf{T}} J = -JB$$

are the complex analogue of J-skew-symmetric matrices (these are the usual skew-Hermitian matrices when $J = E$, see Lecture II.19).

J-unitary matrices make up a group $\mathrm{U}_J(n)$ (a group $\mathrm{U}(n)$ when $J = E$), and J-skew-Hermitian matrices make up a vector space $\mathfrak{u}_J(n)$ (a vector space $\mathfrak{u}(n)$ when $J = E$).

Notice that although the vector space $\mathfrak{u}_J(n)$ consists of complex matrices, it sustains multiplication by only real numbers (after being multiplied by i a skew-Hermitian matrix becomes Hermitian), i.e. $\mathfrak{u}_J(n)$ is *a vector space over the field* $\mathbb{R}$. It is not hard to calculate that for $J = E$ its dimension is n^2. (The condition that the matrices should be skew Hermitian imposes no constraints on $\frac{n(n-1)}{2}$ complex numbers above the main diagonal and requires that n numbers on the main diagonal should be pure imaginary).

Problem 3. Show that for any J-skew-Hermitian matrix B the matrix e^B is a J-unitary matrix and that for a J-unitary matrix $A = e^B$ with $|B| < \ln 2$ the matrix B is necessarily J-skew-Hermitian.

In order to directly generalize the definition we shall also apply the term "matrix Lie group" to groups $\mathcal{G}$ consisting of complex matrices for which there is a vector space $\mathfrak{g} \subset \operatorname{Mat}_n(\mathbb{C})$ (*over the field* $\mathbb{R}$) having Properties 1° and 2° of Definition 1. By this assumption we may say that $\mathrm{U}_J(n)$ is a *matrix Lie group with Lie algebra* $\mathfrak{u}_J(n)$. Since Proposition 1 clearly remains completely valid also for such matrix Lie groups, we see therefore that the *group* $\mathrm{U}_J(n)$ *is a smooth manifold.* When $J = E$ the dimension of this manifold is n^2.

Problem 4. Show that the correspondence

$$\begin{Vmatrix} A_1 & A_2 \\ A_3 & A_4 \end{Vmatrix} \mapsto A_1 + iA_2$$

is a diffeomorphism (and at the same time an isomorphism) of the orthogonal symplectic group $\mathrm{Sp}(m, \mathbb{R}) \cap \mathrm{O}(2m)$ onto a group $\mathrm{U}(m)$.

It follows from relation (19) that for any J-unitary matrix A the equality

$$|\det A| = 1$$

holds.

Unitary matrices A for which

$$\det A = 1$$

(*unimodular unitary matrices*) form a subgroup $\mathrm{SU}(n)$ of the group $\mathrm{U}(n)$. This group is a matrix Lie group with Lie algebra consisting of all trace-free skew-Hermitian n by n matrices (this algebra is usually designated $\mathfrak{su}(n)$). Since clearly $\dim \mathfrak{su}(n) = n^2 - 1$, the *group* $\mathrm{SU}(n)$ *is a smooth manifold of dimension* $n^2 - 1$.

Another interesting subgroup of the group $\mathrm{U}(n)$ is (with even $n = 2m$) a *unitary symplectic group* $\mathrm{Sp}(m)$ consisting of unitary matrices of order $n = 2m$ that satisfy the symplecticity condition (which holds for complex matrices as well). This group is a matrix Lie group with Lie algebra $\mathfrak{sp}(m)$ which consists of skew-Hermitian matrices (17) that satisfy relations (18). But the condition that matrices (17) should be skew-Hermitian reduces to the relations

$$\bar{A}_1 = -A_1, \quad \bar{A}_4^{\top} = -A_4, \quad \bar{A}_2^{\top} = -A_3$$

and these relations together with relations (18) yield

$$A_4 = \bar{A}_1, \quad A_3 = -\bar{A}_2,$$

where A_1 is a skew-Hermitian matrix (which depends on m^2 real parameters), and A_2 is a symmetric matrix (which, being complex, depends on $m(m+1)$ real parameters). This shows that $\dim \mathfrak{sp}(m) = m^2 + m(m+1) = m(2m+1)$ and, consequently, that the *group* Sp (m) *is a smooth manifold of dimension* $m(2m+1)$.

Note that dim Sp (m) = dim Sp $(m, \mathbb{R})$. Nevertheless the *manifolds* Sp (m) and Sp $(m, \mathbb{R})$ *are nondiffeomorphic* (if only one of them is compact and the other is not).

As already noted, the symplecticity condition (and, more generally, the J-orthogonality condition) is also essential for complex matrices. This gives rise to a *group* $\mathrm{O}_J(n, \mathbb{C})$ *of complex J-orthogonal n by n matrices* (and, in particular, to a *group* $\mathrm{O}(n, \mathbb{C})$ *of complex orthogonal n by n matrices* and to a *group* Sp $(m, \mathbb{C})$ *of complex symplectic matrices of order* $n = 2m$). This group satisfies the conditions of Definition 1, the only difference being that the corresponding Lie algebra $\mathfrak{g}$ consists of complex matrices and is a vector space over the field $\mathbb{C}$. Such groups will be called *complex matrix Lie groups*.

Since any vector space over the field $\mathbb{C}$ is trivially a vector space over the field $\mathbb{R}$ (of twice a greater dimension), any complex matrix group is a matrix Lie group (in the above generalized sense) and hence a smooth manifold. In particular, we see that *all groups of the form* $\mathrm{O}_J(n, \mathbb{C})$ (and, in particular, the groups $\mathrm{O}(n, \mathbb{C})$ and Sp $(m, \mathbb{C})$) *are smooth manifolds*, with $\dim \mathrm{O}(n, \mathbb{C}) = n(n-1)$ and dim Sp $(m, \mathbb{C}) = 2m(2m+1)$.

Notice that thus

$$\mathrm{Sp}(m) = \mathrm{Sp}(m, \mathbb{C}) \cap \mathrm{U}(2m)$$

(a similar intersection $\mathrm{Sp}(m, \mathbb{R}) \cap \mathrm{U}(2m)$ is nothing else but the orthogonal symplectic group $\mathrm{Sp}(m, \mathbb{R}) \cap \mathrm{O}(2m)$).

An example of groups $\mathrm{O}_J(n, \mathbb{C})$ suggests an idea of considering in a general form manifolds that have charts (U, h) for which the set $h(U)$ is an open set of the space $\mathbb{C}^n$ (such charts are called *complex*).

Two complex charts (U, h) and (V, k) are said to be *complex compatible* if either $U \cap V = \varnothing$ or the sets $h(U \cap V)$ and $k(U \cap V)$ are open in $\mathbb{C}^n$, and the mapping

$$k \circ h^{-1}: h(U \cap V) \to k(U \cap V)$$

is a *complex analytic diffeomorphism*, i.e. this mapping and its inverse $h \circ k^{-1}$ can be expressed by complex analytic (or, in another terminology, *holomorphic*) functions (which in the neighbourhood of any point $z_0 = (z_0^1, \ldots, z_0^n) \in h(U \cap V) \subset \mathbb{C}^n$ can be expanded in a power series of differences $z^1 - z_0^1, \ldots, z^n - z_0^n$).

A set $\mathscr{X}$ provided with an atlas of complex compatible complex charts is called a *complex analytic manifold*. The number n is called the *complex dimension* of that manifold and denoted by $\dim_{\mathbb{C}} \mathscr{X}$.

The "complex" analogue of Proposition 1 states that *any complex matrix Lie group is a complex analytic manifold*. The proof coincides verbatim with the proof of Proposition 1.

[Note that $\mathrm{U}(n)$ is not a complex matrix Lie group; it is in general impossible to introduce on $\mathrm{U}(n)$ the structure of a complex analytic manifold.]

Since the space $\mathbb{C}^n$ can be naturally identified with space $\mathbb{R}^{2n}$ (it suffices to replace each complex coordinate z^k, $k = 1, \ldots, n$, by its real and imaginary parts) and since for any complex analytic functions of $z^1, \ldots, z^n$ its real and imaginary parts are (prove it!) real analytic functions of

$$x^1 = \mathrm{Re}\, z^1, \ldots, x^n = \mathrm{Re}\, z^n,$$
$$y^1 = \mathrm{Im}\, z^1, \ldots, y^n = \mathrm{Im}\, z^n,$$

every complex analytic manifold $\mathscr{X}$ is trivially a smooth (real analytic) manifold of doubled dimension.

This manifold is called a *realization* of the manifold $\mathscr{X}$ and designated $\mathscr{X}_{\mathrm{R}}$.

Now we must return to general topology.

Definition 2. By a *path* in a topological space $\mathscr{X}$ is meant an arbitrary continuous mapping $u: \mathbf{I} \to \mathscr{X}$ of an interval $\mathbf{I} = [0, 1]$ into $\mathscr{X}$. The point $u(0)$ is called the *initial point* of the path, and $u(1)$ the *terminal point.* It is also said that a path u *joins* in $\mathscr{X}$ a point $p_0 = u(0)$ and a point $p_1 = u(1)$.

Deformations of bases of an n-dimensional vector space $\mathscr{V}$ (see Definition 2 of Lecture I.12) are nothing but paths in Stiefel's manifold $V(n, \mathscr{V}) = \mathrm{GL}(n)$, and similarly defined pseudo-orthonormal deformations of bases of a pseudo-Euclidean space of type (p, q) are nothing but paths in the group $\mathrm{O}(p, q)$.

It is easy to see (cf. Proposition 1 of Lecture I.12) that the relation of "being joined by a path" is an equivalence relation on $\mathscr{X}$, i.e. it is reflexive (every point $p_0 \in \mathscr{X}$ is joined to itself by a *constant path* $u: \mathbf{I} \to \mathscr{X}$ defined by $u(t) = p_0$, $t \in \mathbf{I}$), symmetric (if a path $u: \mathbf{I} \to \mathscr{X}$ joins a point p_0 and p_1, then the *reverse path* $u^{-1}: \mathbf{I} \to \mathscr{X}$ defined by $u^{-1}(t) = u(1 - t)$, $t \in \mathbf{I}$, joins p_1 to p_0), and transitive (if a path $u: \mathbf{I} \to \mathscr{X}$ joins a point p_0 and p_1, and a path $v: \mathbf{I} \to \mathscr{X}$ joins p_1 and p_2, then a path $w: \mathbf{I} \to \mathscr{X}$, defined by

$$(20) \qquad w(t) = \begin{cases} u(2t) & \text{if } 0 \leqslant t \leqslant 1/2, \\ v(2t-1) & \text{if } 1/2 \leqslant t \leqslant 1, \end{cases}$$

and called the *product* uv of paths u and v, joins p_0 and p_2; prove that formula (20) defines a continuous mapping).

Definition 3. Classes in this equivalence relation are called the *pathwise connectedness* of a topological space $\mathscr{X}$. A space $\mathscr{X}$ having only one path-connected component, i.e. such that any two of its points can be joined by a path, is called *pathwise-connected space.*

Orientations of a vector space as classes of bases deformable into each other (see Lecture I.12) and similarly defined orientations of a pseudo-Euclidean space are nothing but pathwise-connected components of groups GL (n) and O (p, q), respectively. We thus see that *groups* GL (n) *and* O (n) *each has two pathwise-connected components.* Similarly it can be shown that the *group* O (p, q) *for* $0 < p < n$ *has four components.* The component of GL (n) that contains the identity of the group (or briefly the *component of the identity*) is the group $\mathrm{GL}^+(n)$ (consisting of a matrix with positive determinant), and the component of the identity of O (n) is the group SO (n). The component of the identity of the group O (p, q) for $0 < p < n$ is designated $\mathrm{O}_+^+(p, q)$.

Problem 5. Prove that Sp $(m, \mathbb{R})$, U (n), SU (n), and Sp (m) are pathwise-connected groups.

There is another notion of connectedness, which is sometimes more convenient (and which we have already used in Lecture 1 in describing the notion of a simple line).

Recall (see Lecture 8) that a subset A of a topological space $\mathscr{X}$ is said to be *closed* if its complement $\mathscr{X} \setminus A$ is open. A subset which is both open and closed is called *open-closed.*

Clearly, the empty set $\varnothing$ and the entire space $\mathscr{X}$ are open-closed. They are called *trivial open-closed subsets.* An example of a nontrivial open-closed subset is a subgroup SO (n) of the group O (n).

Definition 4. A topological space is said to be *connected* if it has no nontrivial open-closed subsets. A subset of a topological space is said to be *connected* if it is connected in the induced topology.

Intuitively the connectedness of a space implies that the space consists of only one piece.

Problem 6. Prove that the Alexandroff half-line (see Remark 1 of Lecture 10) is a connected topological space.

Definition 5. A connected subset A of a topological space $\mathscr{X}$ is said to be its *component* (or more extensively its *connected component*) if it is maximal, i.e. if every connected subset containing A coincides with A.

It is known from calculus that any *closed interval* $[a, b]$ *of the axis* $\mathbb{R}$ *is a connected space.*

We prove this for completeness. Let $[a, b]$ be a disconnected closed interval, i.e. let there be a nonempty open-closed set $C \subset [a, b]$ other than the entire $[a, b]$. Changing, if necessary, to the complement (which is also open-closed) we may assume without loss of generality that $a \in C$. Let T be the set of all points $t \in [a, b]$ for which $[a, t) \subset C$. Since it follows from $[a, t) \subset C$ that $[a, t] \subset C$ (for C is closed), we have $T \subset C$. Since C is open in $[a, b]$ (and $a \in C$) there exists an $\varepsilon > 0$ such that $[a, a+\varepsilon) \subset C$, i.e. $a+\varepsilon \in T$. Hence the set T is not empty, and its upper bound $t_0 = \sup T$ is strictly greater than a. Let $a < t < t_0$. By the definition of the upper bound there is a number $t_1 \in T$ such that $t < t_1$. Then $[a, t) \subset [a, t_1) \subset C$ and hence $t \in T$. This proves that $[a, t_0) \subset C$, i.e. that $t_0 \in T \subset C$. But if $t_0 \in C$ and $t_0 < b$, then, C being open in $[a, b]$, there is an $\varepsilon > 0$ such that $(t_0 - \varepsilon, t_1 + \varepsilon) \subset C$. Therefore $[a, t_0 + \varepsilon) \subset C$ and hence $t_0 + \varepsilon \in T$. Since this contradicts the equation $t_0 = \sup T$, the inequality $t_0 < b$ is impossible. Consequently, $t_0 = b$, and hence, contrary to the hypothesis, $C = [a, b]$. □

In particular, the *closed interval* I *is connected.*

It immediately follows that *any pathwise-connected space* $\mathscr{X}$ *is connected.* Indeed, if C is an arbitrary open-closed set in $\mathscr{X}$, then for every path $u\colon \mathrm{I} \to \mathscr{X}$ the set $u^{-1}C$ of all points $t \in [0, 1]$ for which $u(t) \in C$ is open-closed in I. Hence, the closed interval I being connected, either $u^{-1}C = \varnothing$ or $u^{-1}C = \mathrm{I}$. If $C \neq \varnothing$, then we may exclude the first case choosing the initial point of the path u in C, and if $C \neq \mathscr{X}$, then we may exclude the second case as well, choosing the terminal point of u outside C. Thus, if $\mathscr{X}$ has a nonempty open-closed set C other than $\mathscr{X}$, then it is impossible to join by a path any point of C to any point outside C. Hence $\mathscr{X}$ is not pathwise-connected. □

It is thus clear that *every pathwise-connected component is contained in some (obviously unique) connected component.*

Further, it is easy to see that the *closure* $\bar{A}$ *of an arbitrary connected subset* A *is connected.* Indeed, if a nonempty subset C is open-closed in $\bar{A}$, then its intersection

$C \cap A$ with A is not empty (why?) and open-closed in A. Therefore $C \cap A = A$ and hence $A \subset C$. Since C is closed not only in $\bar{A}$ but also in $\mathcal{X}$, it follows that $\bar{A} \subset C$ and hence $\bar{A} = C$. □

It follows, in particular, that *any component of $\mathcal{X}$ is closed.*

Likewise it can be proved that *if the intersection $\bigcap A_\alpha$ of a family $\{A_\alpha\}$ of connected sets is not empty, then their union $\bigcup A_\alpha$ is connected.* Indeed, if C is open-closed in $\bigcup A_\alpha$ and $C \cap A_\alpha \neq \varnothing$, then $C \cap A_\alpha = A_\alpha$, and hence $A_\alpha \subset C$. But if $A_{\alpha_0} \subset C$ at least for one subscript α_0, then $\bigcap A_\alpha \subset C$, and hence $C \cap A_\alpha \neq \varnothing$ for all α. Therefore either $A_\alpha \subset C$ for all α or $C \cap A_\alpha = \varnothing$ also for all α. In the former case $C = \bigcup A_\alpha$ and in the latter $C = \varnothing$. □

It follows that *for any point $p \in \mathcal{X}$ the component of $\mathcal{X}$ containing the point p is the union of all connected subsets of $\mathcal{X}$ containing p*, and that *the components of $\mathcal{X}$ are mutually disjoint.* If therefore a *space $\mathcal{X}$ has a finite number of components, then every component is open in $\mathcal{X}$* (and is therefore an open-closed set).

To sum up, *an arbitrary topological space $\mathcal{X}$ is the union of mutually disjoint components each of which is a closed (and open if there is a finite number of components) connected subset of the space $\mathcal{X}$.*

A space $\mathcal{X}$ is connected if and only if it has only one component.

A space $\mathcal{X}$ is said to be *totally disconnected* if each of its components consists of only one point. Clearly, a *Hausdorff space $\mathcal{X}$ for which* ind $\mathcal{X} = 0$ (see Lecture 10) *is totally disconnected.*

On the other hand, there are examples (rather complicated) of totally disconnected Hausdorff spaces (even metric spaces and spaces satisfying the second axiom of countability) of arbitrary positive dimension. All these spaces are trivially not compact since it can be proved (try to do it!) that a *totally disconnected Hausdorff compact space $\mathcal{X}$ is zero-dimensional* (ind $\mathcal{X}$ = dim $\mathcal{X}$ = 0).

In contrast to connected components pathwise-connected components are neither open nor closed. But if a space is *locally pathwise-connected* (any point of it has a fundamental system of neighbourhoods each of which is pathwise connected), *then every pathwise-connected component C of the space $\mathscr{X}$ is open and closed.* Indeed, if U is a pathwise connected neighbourhood of an arbitrary point $p \in \bar{C}$, then, the intersection $U \cap C$ being not empty, the union $C \cup U$ is pathwise connected, and hence, since C is maximal, it coincides with C. Thus, $U \subset C$, and therefore $p \in \operatorname{Int} C$. This proves that $\bar{C} \subset \operatorname{Int} C$, and hence that $\bar{C} = \operatorname{Int} C = C$. Consequently, the component C is open-closed. □

It follows that *pathwise-connected components of a locally pathwise connected space coincide with its connected components* (and, in addition, in *a locally pathwise-connected space all its components, whatever their number, are open-closed*).

In particular, a locally pathwise-connected space is connected if and only if it is pathwise-connected.

Since every point of any topological (and in particular any smooth) manifold has a fundamental system of neighbourhoods homeomorphic to an open ball $\mathring{\mathbb{B}}^n$ and since the ball $\mathring{\mathbb{B}}^n$ is obviously pathwise connected, an *arbitrary topological manifold is locally pathwise connected.*

For topological manifolds therefore the components coincide with pathwise-connected components and are open-closed sets. (In particular, *any component of a manifold is itself a manifold.*)

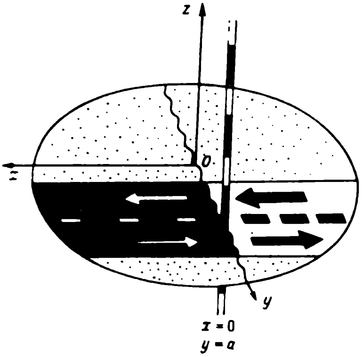
We see that pathwise-connected manifolds are precisely connected manifolds. For this reason the term "pathwise-connected" is not generally applied to manifolds and is replaced by the term "connected".

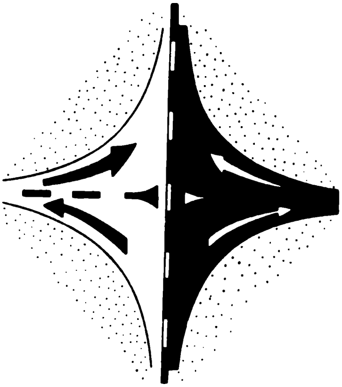
A path $u\colon \mathrm{I} \to \mathscr{X}$ in a smooth manifold $\mathscr{X}$ is said to be *smooth* if it is a restriction on I of some smooth mapping $u'\colon (-\varepsilon, 1 + \varepsilon) \to \mathscr{X}$, where $\varepsilon > 0$. (Recall that the interval $(-\varepsilon, 1 + \varepsilon)$ is a smooth manifold.) If the mapping u' is smooth everywhere except for a finite number of points, the path u is called *piecewise smooth.*

It is clear that if an open set $U \subset \mathscr{X}$ is diffeomorphic to a connected open set in $\mathbb{R}^n$ (to a ball, for example), then any two of its points can be joined in U by a smooth path. It is obvious, since the closed interval $\mathbf{I}$ is compact, that *any two points of a connected smooth manifold can be joined by a piecewise smooth path.*

Problem 7. Prove that any two points of a connected smooth manifold can be joined by even a smooth path.

All examples of connected smooth manifolds in this lecture satisfy the second axiom of countability. There are, however, such connected smooth manifolds that fail to satisfy this axiom. Such a manifold is, say, the Alexandroff half-line (see Problem 7 above). We give a more elementary example which does not use ordinals (this example is due to Calabi and Rosenlicht).





Let $\mathscr{X}$ be a subset of $\mathbb{R}^3$ consisting of points (x, y, z) for which either $x = 0$ or $z = 0$ (the union of coordinate planes Oyz and Oxy), and U_a, $a \in \mathbb{R}$, be a subset consisting of points (x, y, z) for which either $x \neq 0$ or $y = a$ (the union of the plane Oxy, with the ordinate axis $x = 0$ removed, and the straight line $x = 0$, $y = a$). Further, let h_a: $U_a \to \mathbb{R}^2$ be a mapping of the set U_a onto the plane $\mathbb{R}^2$ with coordinates (u_a, v_a) given as follows:

$$u_a = x, \quad v_a = \begin{cases} \dfrac{y-a}{x} & \text{if } x \neq 0, \\ z & \text{if } x = 0. \end{cases}$$

This mapping is obviously bijective (the inverse mapping carries a point $(u_a, v_a) \in \mathbb{R}^2$ into a point $(u_a, a + u_a v_a, v_a) \in U_a$) and hence the pair (U_a, h_a) is a chart on $\mathscr{X}$. For any two such charts (U_a, h_a) and (U_b, h_b) the set $U_a \cap U_b$ is the plane Oxy, with the ordinate axis $x = 0$ removed, the sets $h_a(U_a \cap U_b)$ and $h_b(U_a \cap U_b)$ represent the plane $\mathbb{R}^2$ with the ordinate axis $u = 0$ removed, and the mapping

$$h_b \circ h_a^{-1}: h_a(U_a \cap U_b) \to h_b(U_a \cap U_b)$$

is given by

$$u_b = u_a, \quad v_b = v_a + \frac{a-b}{u_a}$$

and is therefore real analytic. This means that the charts (U_a, h_a), $a \in \mathbb{R}$ are real-analytically compatible and hence define on $\mathscr{X}$ the structure of a two-dimensional smooth C^ω-manifold.

Problem 8. Show that the constructed smooth manifold is Hausdorff, connected, and does not satisfy the second axiom of countability.

This example can be easily extended to greater dimensions. Moreover, if we replace $\mathbb{R}$ by $\mathbb{C}$, then the same construction will yield an example of a two-dimensional connected complex analytic manifold which does not satisfy the second axiom of countability.

It is interesting that for $n = 1$ such an example is impossible since, as shown by the Hungarian mathematician Rado, *every complex curve* (Hausdorff connected one-dimensional complex analytic manifold) *satisfies the second axiom of countability.*

Remark 2. The comparison of definitions shows that complex curves are nothing but **abstract Riemann surfaces** known from the theory of functions of one complex variable.

Lecture 12

Vectors tangent to a smooth manifold · Derivatives of holomorphic functions · Tangent vectors to complex-analytic manifolds · The differential of a smooth mapping · The chain rule · The gradient of a smooth function · The étale mapping theorem · The theorem on the change of local coordinates · Locally flat mappings

---

We introduce the notion of a vector tangent to an arbitrary smooth manifold $\mathscr{X}$.

To understand how this can be done consider a special case of a manifold $\mathbb{R}^n$ when we already know what the vector is.

In arbitrary coordinates $x^1, \ldots, x^n$ (possibly, even curvilinear) a vector $\mathbf{a}$ of an affine space $\mathbb{R}^n$ is given at every point $p_0 \in \mathbb{R}^n$ by n numbers $(a^1, \ldots, a^n)$ (by a vector of a vector space $\mathbb{R}^n$). How are these numbers transformed when changing to other coordinates $x^{1'}, \ldots, x^{n'}$? (From linear algebra we know the answer only in cases where new coordinates can be expressed linearly in terms of the old ones, now we have in mind the most general transformation of coordinates given by arbitrary smooth functions $x^{i'} = x^{i'}(x^1, \ldots, x^n)$, $i' = 1, \ldots, n$.) To give the answer to this question, consider in $\mathbb{R}^n$ an arbitrary smooth curve

$$x^i = x^i(t), \quad i = 1, \ldots, n$$

at $t = t_0$ passing through a point p_0 and having at this point a tangent vector $\mathbf{a}$, i.e. such that $a^i = \dot{x}^i(t_0)$ for any $i = 1, \ldots, n$ (the dot designates differentiation with respect to t). In coordinates $x^{1'}, \ldots, x^{n'}$ this curve is given by functions $x^{i'}(t) = x^{i'}(x^1(t), \ldots, x^n(t))$, $i' = 1', \ldots, n'$, and the tangent vector (i.e. the same vector **a**) has components $a^{i'} = \dot{x}^{i'}(t_0)$. But by the rule for differentiating a composite function

$$\dot{x}^{i'}(t) = \frac{\partial x^{i'}(x^1(t), \ldots, x^n(t))}{\partial x^i} \dot{x}^i(t)$$

(as ever, the summation is taken over i), and hence

$$a^{i'} = \left(\frac{\partial x^{i'}}{\partial x^i}\right)_0 a^i, \tag{1}$$

where $\left(\frac{\partial x^{i'}}{\partial x^i}\right)_0 = \left(\frac{\partial x^{i'}}{\partial x^i}\right)_{p_0}$ are the values of partial derivatives $\frac{\partial x^{i'}}{\partial x^i}$ at a point p_0 with coordinates $x^1(t_0), \ldots, x^n(t_0)$.

Now let $\mathscr{X}$ be an arbitrary (C^r, $r \geqslant 1$) manifold of dimension n. For any point $p_0 \in \mathscr{X}$ by $\mathsf{A}(p_0)$ we denote the set of all charts (U, h) of $\mathscr{X}$ for which $p_0 \in U$ (these charts are said to be *centred* at p_0). The computation we have just made motivates the following definition.

Definition 1. *By a tangent vector to a manifold* $\mathscr{X}$ (or simply a *vector of the manifold* $\mathscr{X}$) at a point $p_0 \in \mathscr{X}$ is meant a mapping

$$A\colon \mathsf{A}(p_0) \to \mathbb{R}^n \tag{2}$$

such that for arbitrary charts $(U, h) = (U, x^1, \ldots, x^n)$ and $(U', h') = (U, x^{1'}, \ldots, x^{n'})$ of $\mathsf{A}(p_0)$ the vectors $A(U, h) = (a^1, \ldots, a^n)$ and $A(U', h') = (a^{1'}, \ldots, a^{n'})$ of the vector space $\mathbb{R}^n$ are connected by formula (1), i.e. by

$$A(U', h') = \left(\frac{\partial h'}{\partial h}\right)_0 A(U, h), \tag{3}$$

where $\left(\frac{\partial h'}{\partial h}\right)_0$ is a linear operator $\mathbb{R}^n \to \mathbb{R}^n$ with matrix $\left\| \left(\frac{\partial x^{i'}}{\partial x^i}\right)_0 \right\|$.

The components $a^1, \ldots, a^n$ of a vector $A(U, h) \in \mathbb{R}^n$ are called the *coordinates of vector A in a chart* (U, h) (or in *local coordinates* $x^1, \ldots, x^n$). To simplify the formulas the equation $A(U, h) = (a^1, \ldots, a^n)$ is usually written as

$$A = (a^1, \ldots, a^n) \text{ in } (U, h).$$

When the chart (U, h) is fixed, the hint "in (U, h)" is as a rule omitted.

The set of all vectors of a manifold $\mathscr{X}$ at a point p_0 is called a *tangent space* of $\mathscr{X}$ at p_0 and designated $\mathbf{T}_{p_0}\mathscr{X}$. It is a vector space over the field $\mathbb{R}$ with respect to linear operations defined as follows:

$$(A + B)(U, h) = A(U, h) + B(U, h),$$
$$(\lambda A)(U, h) = \lambda A(U, h),$$

where $A, B \in \mathbf{T}_{p_0}\mathscr{X}$, $\lambda \in \mathbb{R}$ and $(U, h) \in \mathbf{A}(p)$. [Inclusions $A + B \in \mathbf{T}_{p_0}\mathscr{X}$ and $\lambda A \in \mathbf{T}_{p_0}\mathscr{X}$ are ensured by linear operators $\left(\frac{\partial h'}{\partial h}\right)_0$.]

Thus by definition if $A = (a^1, \ldots, a^n)$ and $B = (b^1, \ldots, b^n)$ in (U, h), then $A + B = (a^1 + b^1, \ldots, a^n + b^n)$ and $\lambda A = (\lambda a^1, \ldots, \lambda a^n)$ in (U, h). For any chart (U, h) therefore the correspondence

$$A \mapsto A(U, h)$$

defines a linear mapping

$$\mathbf{T}_{p_0}\mathscr{X} \to \mathbb{R}^n. \tag{4}$$

It is easy to see that *mapping (4) is an isomorphism*. Indeed, if $A(U, h) = 0$, then in view of formula (3) $A(U', h') = 0$ and for every chart $(U', h') \in \mathbf{A}(p_0)$, i.e. $A = 0$. This means that mapping (4) is monomorphic. On the other hand, putting for any vector $\mathbf{a} = (a^1, \ldots, a^n) \in \mathbb{R}^n$ and any chart $(U', h') \in \mathbf{A}(p_0)$

$$A(U', h') = \left(\frac{\partial h'}{\partial h}\right)_0 \mathbf{a},$$

we obtain a mapping $A\colon \mathbf{A}(p_0) \to \mathbb{R}^n$ having the property that $A(U, h) = \mathbf{a}$. So to prove that mapping (4) is an epimorphism (and hence an isomorphism) it suffices to show that $A \in \mathbf{T}_{p_0}\mathscr{X}$, i.e. that

$$A(U'', h'') = \left(\frac{\partial h''}{\partial h'}\right)_0 A(U', h')$$

for any two charts (U', h'), $(U'', h'') \in \mathsf{A}(p_0)$. But by definition:

$$A(U'', h'') = \left(\frac{\partial h''}{\partial h}\right)_0 \mathbf{a} \text{ and } A(U', h') = \left(\frac{\partial h'}{\partial h}\right)_0 \mathbf{a},$$

and by the chain rule on the differentiation of composite functions

$$\left(\frac{\partial x^{i''}}{\partial x^i}\right)_0 = \left(\frac{\partial x^{i''}}{\partial x^{i'}}\right)_0 \left(\frac{\partial x^{i'}}{\partial x^i}\right)_0,$$

i.e. $\left(\frac{\partial h''}{\partial h}\right)_0 = \left(\frac{\partial h''}{\partial h'}\right)_0 \left(\frac{\partial h'}{\partial h}\right)_0$. Hence

$$A(U'', h'') = \left(\frac{\partial h''}{\partial h}\right)_0 \mathbf{a} = \left(\frac{\partial h''}{\partial h'}\right)_0 A(U', h'). \qquad \Box$$

Thus $\mathbf{T}_{p_0}\mathscr{X}$ *is a vector space of dimension* n:

$$\dim \mathbf{T}_{p_0}\mathscr{X} = \dim \mathscr{X}.$$

Vectors transforming under isomorphism (4) to the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of the space $\mathbb{R}^n$ ase designated

$$\left(\frac{\partial}{\partial x^1}\right)_{p_0}, \ldots, \left(\frac{\partial}{\partial x^n}\right)_{p_0}. \tag{5}$$

They make up a basis for $\mathbf{T}_{p_0}\mathscr{X}$, with the coordinates of a vector $A \in \mathbf{T}_{p_0}\mathscr{X}$ with respect to this basis being precisely its coordinates in a chart (U, h): if $A = (a^1, \ldots, a^n)$ in (U, h), then:

$$A = a^i \left(\frac{\partial}{\partial x^i}\right) p_0 \tag{6}$$

and vice versa.

When $\mathscr{X}$ is the space $\mathbb{R}^n$, among all isomorphisms (4) there is one selected isomorphism corresponding to the chart $(\mathbb{R}^n, \mathrm{id})$, and via this isomorphism we can identify the space $\mathbf{T}_{p_0}\mathbb{R}^n$ with $\mathbb{R}^n$. Thus

$$\mathbf{T}_{p_0}\mathbb{R}^n = \mathbb{R}^n \tag{7}$$

for any point $p_0 \in \mathbb{R}^n$, which is in full agreement with what we have proceeded.

Since for any point p_0 of any open submanifold U of an arbitrary manifold $\mathscr{X}$ the space $\mathbf{T}_{p_0}U$ can be naturally identified with $\mathbf{T}_{p_0}\mathscr{X}$:

$$\mathbf{T}_{p_0}U = \mathbf{T}_{p_0}\mathscr{X},$$

we get, in particular,

$$\mathbf{T}_{p_0}U = \mathbb{R}^n \tag{8}$$

for any point $p_0 \in U$ of an arbitrary open set $U \subset \mathbb{R}^n$.

We shall constantly use these identifications, though not always explicitly.

More generally, if $\mathscr{X}$ is a vector space $\mathscr{V}$ (or an open set in $\mathscr{V}$), then among isomorphisms (4) there are isomorphisms corresponding to charts of the form $(\mathscr{V}, h)$, where h is a coordinate isomorphism $\mathscr{V} \to \mathbb{R}^n$ associated with some basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in $\mathscr{V}$. Since for two such isomorphisms h and h' the linear operator $\left(\frac{\partial h'}{\partial h}\right)_0$, $p_0 \in \mathscr{V}$, obviously coincides with the operator $h' \circ h^{-1}$ (and is, in particular, independent of p_0), the composition $\mathbf{T}_{p_0}\mathscr{X} \to \mathscr{V}$ of isomorphism (4) corresponding to a chart $(\mathscr{V}, h)$ and an isomorphism h^{-1}: $\mathbb{R}^n \to \mathscr{V}$ is the same for all h. Thus $\mathbf{T}_{p_0}\mathscr{V}$ *can be naturally identified with*

$$\mathbf{T}_{p_0}\mathscr{V} = \mathscr{V} \text{ for any } p_0 \in \mathscr{V}.$$

Notice that in this identification it is a basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ that corresponds to basis (5).

If, however, $\mathscr{X}$ is an affine space $\mathscr{A}$, then for any point $p_0 \in \mathscr{A}$ the tangent space $\mathbf{T}_{p_0}\mathscr{A}$ is similarly identified with the associated vector space $\mathscr{V}$ (and is therefore the same for all p_0). At this point it is, however, convenient to forcibly introduce the dependence on p_0 and identify $\mathbf{T}_{p_0}\mathscr{A}$ with the space $\mathscr{A}$ in which p_0 is taken as the reference point, i.e. to assume all vectors of $\mathbf{T}_{p_0}\mathscr{A}$ to be marked off from p_0.

When $\mathscr{X}$ is an elementary surface in an affine space $\mathscr{A}$, the tangent space $\mathbf{T}_{p_0}\mathscr{X}$ can be naturally identified with the tangent space of Lecture 3 (which is a subspace of the associated vector space $\mathscr{V}$). Namely, if u, v are local coordinates on $\mathscr{X}$ corresponding to some parametrization $\mathbf{r} = \mathbf{r}(u, v)$ of a surface $\mathscr{X}$, then the corresponding basis vectors $\left(\frac{\partial}{\partial u}\right)_{p_0}$, $\left(\frac{\partial}{\partial v}\right)_{p_0}$ of $\mathbf{T}_{p_0}\mathscr{X}$ can be identified with vectors $\mathbf{r}_{u_0}$, $\mathbf{r}_{v_0}$ of $\mathscr{V}$. Since when the coordinates are changed the vectors $\left(\frac{\partial}{\partial u}\right)_{p_0}$, $\left(\frac{\partial}{\partial v}\right)_{p_0}$ are transformed by the same formulas as the vectors $\mathbf{r}_{u0}$, $\mathbf{r}_{v0}$ this identification of the space $\mathbf{T}_{p_0}\mathscr{X}$ with a subspace of $\mathscr{V}$ is independent of the choice of parametrization $\mathbf{r} = \mathbf{r}(u, v)$.

The notion of a tangent space can immediately be reduced to the case of complex analytic manifolds (see Lecture 11). To realize this we need some notation from complex analysis, which we shall recall in the first place.

For an arbitrary complex function $w(z) = u(x, y) + iv(x, y)$ of a complex variable $z = x + iy$ we set

$$\frac{\partial w}{\partial z} = \frac{1}{2}\left(\frac{\partial w}{\partial x} - i\frac{\partial w}{\partial y}\right) = \frac{1}{2}\left[\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + i\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)\right] \tag{9}$$

and

$$\frac{\partial w}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial w}{\partial x} + i\frac{\partial w}{\partial y}\right) = \frac{1}{2}\left[\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right) + i\left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\right)\right].$$

[The basis for this notation is the formula

$$dw = \frac{\partial w}{\partial z}\,dz + \frac{\partial w}{\partial \bar{z}}\,d\bar{z},$$

where $dz = dx + i\,dy$ and $d\bar{z} = dx - i\,dy$.]

The Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$$

which characterize holomorphic functions can now be written as one equation

$$\frac{\partial w}{\partial \bar{z}} = 0.$$

[Holomorphic functions can thus be interpreted as functions "independent of $\bar{z}$"! For this reason arbitrary functions of z are sometimes written as $w = w(z, \bar{z})$, the notation $w = w(z)$ being used only for holomorphic functions.]

Notice that if a function $w = w(z)$ is holomorphic, then by virtue of the Cauchy-Riemann relations

$$\frac{\partial w}{\partial z} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x},$$

i.e. *for a holomorphic function* $w = w(z)$ *the derivative* $\frac{\partial w}{\partial z}$ *with respect to* z *coincides with the usual derivative* $\frac{\partial w}{\partial x}$ *with respect to* x. Similarly,

$$\frac{\partial w}{\partial z} = \frac{1}{i}\,\frac{\partial w}{\partial y}.$$

It follows for any holomorphic functions $w = w(z)$ and $\zeta = \zeta(w)$ that the standard formula

$$\frac{\partial \zeta}{\partial x} = \frac{\partial \zeta}{\partial u}\,\frac{\partial u}{\partial x} + \frac{\partial \zeta}{\partial v}\,\frac{\partial v}{\partial x}$$

for an arbitrary composite function $\zeta = \zeta(w(z))$ can be rewritten as

$$\frac{\partial \zeta}{\partial z} = \frac{\partial \zeta}{\partial w}\,\frac{\partial w}{\partial z}.$$

This means that *for derivatives* (9) *of holomorphic functions the usual chain rule is valid.*

For complex functions $w(z) = w(z^1, \ldots, z^n)$ of n complex variables partial derivatives $\frac{\partial w}{\partial z^j}$ and $\frac{\partial w}{\partial \bar{z}^j}$, $1 \leqslant j \leqslant n$, can be similarly defined, and a function w is holomorphic if and only if

$$\frac{\partial w}{\partial \bar{z}^1} = 0, \ \ldots, \ \frac{\partial w}{\partial \bar{z}^n} = 0.$$

Every mapping $w\colon U \to \mathbb{C}^m$ where $U \subset \mathbb{C}^n$, is given by m complex functions

(10) $\quad w^1 = w^1(z), \ldots, w^m = w^m(z), \quad z = (z^1, \ldots, z^m) \in U,$

of n complex variables $z^1, \ldots, z^n$ defined on U. Identifying $\mathbb{C}^n$ with $\mathbb{R}^{2n} = \mathbb{R}^n \oplus \mathbb{R}^n$ via the correspondence

$$z = (z^1, \ldots, z^n) \Longleftrightarrow (x^1, \ldots, x^n, y^1, \ldots, y^n) = (\mathbf{x}, \mathbf{y})$$

and $\mathbb{C}^m$ with $\mathbb{R}^{2m} = \mathbb{R}^m \oplus \mathbb{R}^m$ via the correspondence

$$\mathbf{w} = (w^1, \ldots, w^n) \Longleftrightarrow (u^1, \ldots, u^m, v^1, \ldots, v^m) = (\mathbf{u}, \mathbf{v}),$$

where

$$x^1 = \operatorname{Re} z^1, \ldots, x^n = \operatorname{Re} z^n, \quad u^1 = \operatorname{Re} w^1, \ldots, u^m = \operatorname{Re} w^m,$$

$$y^1 = \operatorname{Im} z^1, \ldots, y^n = \operatorname{Im} z^n, \quad v^1 = \operatorname{Im} w^1, \ldots, v^m = \operatorname{Im} w^m,$$

we may think of w as a mapping of $U \subset \mathbb{R}^{2n}$ into $\mathbb{R}^{2m}$ given by $2m$ real functions

$$u^j = u^j(\mathbf{x}, \mathbf{y}), \quad v^j = v^j(\mathbf{x}, \mathbf{y}), \quad 1 \leqslant j \leqslant n.$$

If this mapping is smooth, then its *Jacobian matrix* (matrix of partial derivatives) is therefore of the form

$$\left\| \begin{matrix} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} & \frac{\partial \mathbf{u}}{\partial \mathbf{y}} \\ \frac{\partial \mathbf{v}}{\partial \mathbf{x}} & \frac{\partial \mathbf{v}}{\partial \mathbf{y}} \end{matrix} \right\|,$$

where

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \left\| \frac{\partial u^j}{\partial x^k} \right\|, \quad \frac{\partial \mathbf{u}}{\partial \mathbf{y}} = \left\| \frac{\partial u^j}{\partial y^k} \right\|, \quad \frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \left\| \frac{\partial v^j}{\partial x^k} \right\|, \quad \frac{\partial \mathbf{v}}{\partial \mathbf{y}} = \left\| \frac{\partial v^j}{\partial y^k} \right\|.$$

Therefore

$$\left\| \begin{matrix} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} & \frac{\partial \mathbf{u}}{\partial \mathbf{y}} \\ \frac{\partial \mathbf{v}}{\partial \mathbf{x}} & \frac{\partial \mathbf{v}}{\partial \mathbf{y}} \end{matrix} \right\| = \left\| \begin{matrix} \frac{E_n}{2} & \frac{E_n}{2} \\ \frac{E_n}{2} & -\frac{E_n}{2} \end{matrix} \right\| \cdot \left\| \begin{matrix} \frac{\partial \mathbf{w}}{\partial \mathbf{z}} & \frac{\partial \mathbf{w}}{\partial \bar{\mathbf{z}}} \\ \frac{\partial \bar{\mathbf{w}}}{\partial \bar{\mathbf{z}}} & \frac{\partial \bar{\mathbf{w}}}{\partial \mathbf{z}} \end{matrix} \right\| \cdot \left\| \begin{matrix} E_m & iE_m \\ E_m & -iE_m \end{matrix} \right\|,$$

where E_n and E_m are unit matrices of order n and m, respectively, and

$$\frac{\partial \mathbf{w}}{\partial \mathbf{z}} = \left\| \frac{\partial w^j}{\partial z^k} \right\|, \quad \frac{\partial \mathbf{w}}{\partial \bar{\mathbf{z}}} = \left\| \frac{\partial w^j}{\partial \bar{z}^k} \right\|.$$

When a *mapping w is holomorphic* $\left(\text{i.e. functions (10) are holomorphic and hence } \frac{\partial \mathbf{w}}{\partial \bar{\mathbf{z}}} = 0\right)$, it follows that

$$\left\| \begin{matrix} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} & \frac{\partial \mathbf{u}}{\partial \mathbf{y}} \\ \frac{\partial \mathbf{v}}{\partial \mathbf{x}} & \frac{\partial \mathbf{v}}{\partial \mathbf{y}} \end{matrix} \right\| = \left\| \begin{matrix} \frac{E_n}{2} & \frac{E_n}{2} \\ \frac{E_n}{2i} & -\frac{E_n}{2i} \end{matrix} \right\| \times \tag{11}$$

$$\times \left\| \begin{matrix} \frac{\partial \mathbf{w}}{\partial \mathbf{z}} & 0 \\ 0 & \frac{\partial \bar{\mathbf{w}}}{\mathbf{z}} \end{matrix} \right\| \cdot \left\| \begin{matrix} E_m & iE_m \\ E_m & -iE_m \end{matrix} \right\|.$$

For $n = m$ it follows that *the Jacobian D_w of a holomorphic mapping w: $U \to \mathbb{C}^n$ is expressed by the formula*

$$D_w = \left\| \frac{\partial \mathbf{w}}{\partial \mathbf{z}} \right\|^2$$

and is therefore positive.

Moreover, since

$$\begin{Vmatrix} E & iE \\ E & -iE \end{Vmatrix} \cdot \begin{Vmatrix} \dfrac{E}{2} & \dfrac{E}{2} \\ \dfrac{E}{2i} & -\dfrac{E}{2i} \end{Vmatrix} = \begin{Vmatrix} E & 0 \\ 0 & E \end{Vmatrix},$$

it also follows (for any n and m) that when composing mappings their matrices

$$\begin{Vmatrix} \dfrac{\partial \mathbf{w}}{\partial \mathbf{z}} & \dfrac{\partial \mathbf{w}}{\partial \bar{\mathbf{z}}} \\ \dfrac{\overline{\partial \mathbf{w}}}{\partial \bar{\mathbf{z}}} & \dfrac{\overline{\partial \mathbf{w}}}{\partial \mathbf{z}} \end{Vmatrix}$$

are multiplied (for this is true for Jacobian matrices).

It follows for holomorphic mappings that *when composing holomorphic mappings their matrices* $\dfrac{\partial \mathbf{w}}{\partial \mathbf{z}}$ *are multiplied*: if $w^k = w^k(\mathbf{z})$ and $\zeta^l = \zeta^l(\mathbf{w})$, then

$$\frac{\partial \zeta^l}{\partial z^j} = \frac{\partial \zeta^l}{\partial w^k} \cdot \frac{\partial w^k}{\partial z^j}$$

(**the chain rule for holomorphic mappings**).

We thus see that *for holomorphic functions derivatives* (9) *obey the same formal rules as do the usual partial derivatives of smooth real functions.*

Now let $\mathscr{X}$ be a complex-analytic manifold and let $p_0 \in \mathscr{X}$. Also let $\mathsf{A}(p_0)$ be the set of all complex charts (U, h) of $\mathscr{X}$ for which $p_0 \in U$. For any charts $(U, h) = (U, z^1, \ldots, z^n)$, $(U', h') = (U', z^{1'}, \ldots, z^{n'})$ of $\mathsf{A}(p_0)$ by definition the mapping $h' \circ h^{-1}$ is given by holomorphic functions

(12) $$z^{1'} = z^{1'}(\mathbf{z}), \ldots, z^{n'} = z^{n'}(\mathbf{z}), \text{ where } \mathbf{z} = (z^1, \ldots, z^n).$$

Let $\left(\dfrac{\partial h'}{\partial h}\right)_0$ be a linear operator $\mathbb{C}^n \to \mathbb{C}^n$ with matrix

$$\left(\frac{\partial \mathbf{z}'}{\partial \mathbf{z}}\right)_0 = \left\Vert \left(\frac{\partial z^{j'}}{\partial z^j}\right)_0 \right\Vert, \quad j, j' = 1, \ldots, n,$$

where $\left(\frac{\partial z^{j'}}{\partial z^j}\right)_0$ are values of the derivatives $\frac{\partial z^{j'}}{\partial z^j}$ of functions (12) at a point p_0. By analogy with Definition 1 we shall mean by the *tangent vector* to a manifold $\mathscr{X}$ at a point p_0 a mapping

$$C\colon \mathsf{A}(p_0) \to \mathbb{C}^n \tag{13}$$

such that for any charts (U, h) and (U', h') of $\mathsf{A}(p_0)$

$$C(U', h') = \left(\frac{\partial h'}{\partial h}\right)_0 C(U, h). \tag{14}$$

Such vectors make up a vector space $\mathbf{T}_{p_0}\mathscr{X}$ over a field $\mathbb{C}$ of dimension $n = \dim_{\mathbb{C}} \mathscr{X}$, the mapping

$$\mathbf{T}_{p_0}\mathscr{X} \to \mathbb{C}^n, \quad C \mapsto C(U, h) \tag{15}$$

being an isomorphism for any chart (U, h) of $\mathsf{A}(p_0)$. Vectors transformed under isomorphism (15) into the standard basis for the space $\mathbb{C}^n$ are designated

$$\left(\frac{\partial}{\partial z^1}\right)_{p_0}, \ldots, \left(\frac{\partial}{\partial z^n}\right)_{p_0}.$$

They make a basis for $\mathbf{T}_{p_0}\mathscr{X}$ having the property that $C(U, h) = (c^1, \ldots, c^n)$ is equivalent to $C = c^j \left(\frac{\partial}{\partial z^j}\right)_{p_0}$.

It is interesting to compare a complex n-dimensional space $\mathbf{T}_{p_0}\mathscr{X}$ with a real $2n$-dimensional space $\mathbf{T}_{p_0}\mathscr{X}_{\mathbb{R}}$, where $\mathscr{X}_{\mathbb{R}}$ is a realization of a manifold $\mathscr{X}$ (see Lecture 11).

For any complex chart (U, h) a mapping h considered as a mapping in $\mathbb{R}^{2n}$ is denoted by $h_{\mathbb{R}}$ and the set of all charts of the form $(U, h_{\mathbb{R}})$, where $(U, h) \in \mathsf{A}(p_0)$ by $\mathsf{A}(p_0)_{\mathbb{R}}$. By definition the set $\mathsf{A}(p_0)_{\mathbb{R}}$ is contained in the set $\mathsf{A}_{\mathbb{R}}(p_0)$ of all charts of $\mathscr{X}_{\mathbb{R}}$ at p_0. Although $\mathsf{A}(p_0)_{\mathbb{R}} \neq \mathsf{A}_{\mathbb{R}}(p_0)$, the operation of restricting mappings $\mathsf{A}_{\mathbb{R}}(p_0) \to \mathbb{R}^{2n}$ to $\mathsf{A}(p_0)_{\mathbb{R}}$ clearly allows us to identify every tangent vector $\mathsf{A}_{\mathbb{R}}(p_0) \to \mathbb{R}^{2n}$ of $\mathscr{X}_{\mathbb{R}}$ at p_0 with the mappings

$$A\colon \mathsf{A}(p_0)_{\mathbb{R}} \to \mathbb{R}^{2n} \tag{16}$$

satisfying for any chart (U, h) and (U', h') of $\mathbf{A}(p_0)$ the relation

$$A(U', h'_{\mathrm{R}}) = \left(\frac{\partial h'_{\mathrm{R}}}{\partial h_{\mathrm{R}}}\right)_0 A(U, h_{\mathrm{R}}). \tag{17}$$

On the other hand, by virtue of identification $\mathbb{C}^n = \mathbb{R}^{2n}$ every tangent vector (13) of $\mathscr{X}$ at p_0 can be considered as a mapping $C_{\mathrm{R}}\colon \mathbf{A}(p_0)_{\mathrm{R}} \to \mathbb{R}^{2n}$. For any charts (U, h) and (U', h') of $\mathbf{A}(p_0)$ we have

$$C_{\mathrm{R}}(U', h'_{\mathrm{R}}) = \left(\frac{\partial h'}{\partial h}\right)_0^{\mathrm{R}} C_{\mathrm{R}}(U, h), \tag{18}$$

where $\left(\frac{\partial h'}{\partial h}\right)_0^{\mathrm{R}}$ is the operator $\left(\frac{\partial h'}{\partial h}\right)_0$ considered as the operator $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$.

But it is clear that if some linear operator $\mathbb{C}^n \to \mathbb{C}^n$ is given (in the standard basis) by a matrix $C = A + iB$, then the same operator considered as the operator $\mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is given by the matrix

$$\begin{Vmatrix} A & -B \\ B & A \end{Vmatrix}.$$

Since the role of a matrix of the operator $\left(\frac{\partial h}{\partial h}\right)_0$ is played by the matrix

$$\left(\frac{\partial \mathbf{z}'}{\partial \mathbf{z}}\right)_0 = \left(\frac{\partial \mathbf{x}'}{\partial \mathbf{x}}\right)_0 + i\left(\frac{\partial \mathbf{y}'}{\partial \mathbf{x}}\right)_0,$$

it follows that the operator $\left(\frac{\partial h'}{\partial h}\right)_0^{\mathrm{R}}$ is given by the matrix

$$\begin{Vmatrix} \left(\frac{\partial \mathbf{x}'}{\partial \mathbf{x}}\right)_0 & -\left(\frac{\partial \mathbf{y}'}{\partial \mathbf{x}}\right)_0 \\ \left(\frac{\partial \mathbf{y}'}{\partial \mathbf{x}}\right)_0 & \left(\frac{\partial \mathbf{x}'}{\partial \mathbf{x}}\right)_0 \end{Vmatrix}$$

which by virtue of the Cauchy-Riemann relations coincides with the matrix

$$\begin{Vmatrix} \left(\frac{\partial \mathbf{x}'}{\partial \mathbf{x}}\right)_0 & \left(\frac{\partial \mathbf{x}'}{\partial \mathbf{y}}\right)_0 \\ \left(\frac{\partial \mathbf{y}'}{\partial \mathbf{x}}\right)_0 & \left(\frac{\partial \mathbf{y}'}{\partial \mathbf{y}}\right)_0 \end{Vmatrix}$$

of the operator $\left(\frac{\partial h'_{\mathbb{R}}}{\partial h_{\mathbb{R}}}\right)_0$. This shows that

$$\left(\frac{\partial h'}{\partial h}\right)_0^{\mathbb{R}} = \left(\frac{\partial h'_{\mathbb{R}}}{\partial h_{\mathbb{R}}}\right)_0,$$

and hence formula (18) coincides with (17) (for $A = C_{\mathbb{R}}$). Consequently, $C_{\mathbb{R}}$ is a tangent vector to $\mathscr{X}_{\mathbb{R}}$ at p_0.

The constructed mapping

$$\mathbf{T}_{p_0}\mathscr{X} \to \mathbf{T}_{p_0}\mathscr{X}_{\mathbb{R}}, \quad C \mapsto C_{\mathbb{R}}$$

is obviously bijective and preserves the sums and products of real numbers. In other words, it is an isomorphism of the space $\mathbf{T}_{p_0}\mathscr{X}$, which by virtue of the inclusion $\mathbb{R} \subset \mathbb{C}$ is considered to be a vector space over the field $\mathbb{R}$, onto the space $\mathbf{T}_{p_0}\mathscr{X}_{\mathbb{R}}$. In the natural notation this means that *there is a natural identification*

$$(\mathbf{T}_{p_0}\mathscr{X})_{\mathbb{R}} = \mathbf{T}_{p_0}\mathscr{X}_{\mathbb{R}}. \tag{19}$$

In this sense the manifolds $\mathscr{X}$ and $\mathscr{X}_{\mathbb{R}}$ have the same tangent vectors.

The basis for $\mathbf{T}_{p_0}\mathscr{X}$

$$\left(\frac{\partial}{\partial z^1}\right)_{p_0}, \ldots, \left(\frac{\partial}{\partial z^n}\right)_{p_0}$$

corresponding to a chart (U, h) of $\mathbf{A}(p_0)$ generates the basis

$$\left(\frac{\partial}{\partial z^1}\right)_{p_0}, \ldots, \left(\frac{\partial}{\partial z^n}\right)_{p_0}, J\left(\frac{\partial}{\partial z^1}\right)_{p_0}, \ldots, J\left(\frac{\partial}{\partial z^n}\right)_{p_0} \tag{20}$$

for the space $(\mathbf{T}_{p_0}\mathscr{X})_{\mathbb{R}}$, where J is the operator of a complex structure (of multiplying by i). It is easy to see that by virtue of *identification* (19) *basis* (20) *coincides with the basis*

$$\left(\frac{\partial}{\partial x^1}\right)_{p_0}, \ldots, \left(\frac{\partial}{\partial x^n}\right)_{p_0}, \left(\frac{\partial}{\partial y^1}\right)_{p_0}, \ldots, \left(\frac{\partial}{\partial y^n}\right)_{p_0}$$

for $\mathbf{T}_{p_0}\mathscr{X}_{\mathbb{R}}$ *complying with the chart* $(U, h_{\mathbb{R}})$.

In what follows complex-analytic manifolds are no longer considered. Carrying over, when possible, the results proved for smooth manifolds to the case of complex analytic manifolds is thus left to the reader.

Let $\mathscr{X}$ and $\mathscr{Y}$ be two smooth manifolds (of dimensions n and m, respectively), and let $f\colon \mathscr{X} \to \mathscr{Y}$ be an arbitrary smooth mapping.

Also let p be an arbitrary point of $\mathscr{X}$, $q = f(p)$ its image in $\mathscr{Y}$, and

$$(U, h) = (U, x^1, \ldots, x^n) \text{ and } (V, k) = (V, y^1, \ldots, y^m)$$

be charts of $\mathscr{X}$ and $\mathscr{Y}$ respectively, such that $p \in U$ and $fU \subset Y$.

Then, as we already know (see Lecture 7), the mapping f can be written in charts (U, h) and (V, k) by formulas of the form

$$y^j = f^j(x^1, \ldots, x^n), \quad j = 1, \ldots, m,$$

where f^j are some smooth functions.

An n by m matrix

$$\left\| \left(\frac{\partial f^j}{\partial x^i} \right)_p \right\|, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m \tag{21}$$

whose elements are the values $\left(\frac{\partial f^j}{\partial x^i}\right)_p$ of partial derivatives of functions f^j with respect to x^i at a point p (i.e. more exactly at a point $h(p) \in \mathbb{R}^n$) is called a *Jacobian matrix* of a mapping f in charts (U, h) and (V, k).

This matrix gives a linear mapping $\mathbb{R}^n \to \mathbb{R}^m$ carrying a vector $(a^1, \ldots, a^n) \in \mathbb{R}^n$ into a vector $(b^1, \ldots, b^m) \in \mathbb{R}^m$, where

$$b^j = \left(\frac{\partial f^j}{\partial x^i} \right)_p a^i, \quad j = 1, \ldots, m. \tag{22}$$

Coordinate isomorphisms $A \mapsto A(U, h)$ and $A \mapsto A(V, k)$ allow us to interpret this mapping as a linear mapping $\mathbf{T}_p\mathscr{X} \to \mathbf{T}_q\mathscr{Y}$ of tangent spaces.

Definition 2. The constructed mapping $\mathbf{T}_p\mathscr{X} \to \mathbf{T}_q\mathscr{Y}$ is called a *differential of a smooth mapping* f at a point p. It is designated $(df)_p$ or simply df_p.

For elementary surfaces this construction has already been given in Lecture 3.

Thus if $A = (a^1, \ldots, a^n)$ in a chart (U, h), then $df_p(A) = (b^1, \ldots, b^m)$ in a chart (V, k), where b^j, $j = 1, \ldots, m$, are numbers (22). In linear-algebraic terms this means that the *mapping* df_p *is nothing but a linear mapping*

$\mathbf{T}_p\mathcal{X} \to \mathbf{T}_q\mathcal{Y}$ *with matrix* (21) *in bases* $\left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^n}\right)_p$ *and* $\left(\frac{\partial}{\partial y^1}\right)_q, \ldots, \left(\frac{\partial}{\partial y^m}\right)_p$.

It is, of course, necessary to verify that this definition is correct, i.e. that the mapping df_p is independent of the choice of charts (U, h) and (V, k).

Let $(U', h') = (U', x^{1'}, \ldots, x^{n'})$ and $(V', k') = (V', y^{1'}, \ldots, y^{n'})$ be other charts of $\mathcal{X}$ and $\mathcal{Y}$ having the property that $p \in U'$ and $q \in V'$, and let $d'f_p$ be a mapping df_p constructed using these charts. Then if $A = (a^{1'}, \ldots, a^{n'})$ in (U', h'), then $d'f_p(A) = (b^{1'}, \ldots, b^{m'})$, where

$b^{j'} = \left(\frac{\partial f^{j'}}{\partial x^{i'}}\right)_p a^{i'}$ and $\left(\frac{\partial f^{j'}}{\partial x^{i'}}\right)_p$ are values at a point p for partial derivatives with respect to $x^{i'}$ of functions $f^{j'}$ that express the mapping f in the charts (U', h') and (V', k').

Further, let

$$x^i = x^i(x^{1'}, \ldots, x^{n'}) \text{ and } y^{j'} = y^{j'}(y^1, \ldots, y^m)$$

be expressions for local coordinates x^i and $y^{j'}$ in terms of local coordinates $x^{i'}$ and y^j.

To simplify the formulas we put

$$\mathbf{x} = (x^1, \ldots, x^n) \text{ (and similarly } \mathbf{x}' = (x^{1'}, \ldots, x^{n'})),$$
$$\mathbf{y} = (y', \ldots, y^m) \text{ (and similarly } \mathbf{y}' = (\mathbf{y}^{1'}, \ldots, y^{m'})).$$

Accordingly we shall write $x^i(\mathbf{x}')$ instead of $x^i(x^{1'}, \ldots, x^{n'})$ and $x(\mathbf{x}')$ instead of $(x^1(\mathbf{x}), \ldots, x^n(\mathbf{x}))$. We shall use $\mathbf{y}'(\mathbf{y})$, $\mathbf{f}(\mathbf{x})$, and $\mathbf{f}'(\mathbf{x}')$ in a similar meaning.

In this notation functions f^j and $f^{j'}$ are related as follows:

$$\mathbf{f}'(\mathbf{x}') = \mathbf{y}'(\mathbf{f}(\mathbf{x}(\mathbf{x}')))$$

from which it follows by the rule for differentiating a composite function that

$$\left(\frac{\partial f^{j'}}{\partial x^{i'}}\right)_p = \left(\frac{\partial y^{j'}}{\partial y^j}\right)_q \left(\frac{\partial f^j}{\partial x^i}\right)_p \left(\frac{\partial x^i}{\partial x^{i'}}\right)_p.$$

Since

$$a^i = \left(\frac{\partial x^i}{\partial x^{i'}}\right)_p a^{i'} \text{ and } b^j = \left(\frac{\partial f^j}{\partial x^i}\right)_p a^i,$$

it follows that

$$b^{j'} = \left(\frac{\partial f^{j'}}{\partial x^{i'}}\right)_p a^{i'} = \left(\frac{\partial y^{j'}}{\partial y^j}\right)_q \left(\frac{\partial f^j}{\partial x^i}\right)_p \left(\frac{\partial x^i}{\partial x^{i'}}\right)_p a^{i'}$$

$$= \left(\frac{\partial y^{j'}}{\partial y^j}\right)_q \left(\frac{\partial f^j}{\partial x^i}\right)_p a^i = \left(\frac{\partial y^{j'}}{\partial y^j}\right)_q b^j,$$

i.e. the numbers $b^{j'}$ and b^j are the coordinates (in charts (V', k') and (V, k)) of the same tangent vector. Consequently, $d'f_p(A) = df_p(A)$ for any vector $A \in \mathbf{T}_p\mathscr{X}$, and hence $d'f_p = df_p$. □

Cf. the corresponding reasoning in Lecture 3.

Notice that the mapping df_p only depends on the local behaviour of the mapping f in the neighbourhood of the point p, i.e. *if for mappings f, g: $\mathscr{X} \to \mathscr{Y}$ there exists a neighbourhood of a point p such that $f = g$ on U, then $df_p = dg_p$.*

If $\mathscr{X}$ and $\mathscr{Y}$ are open submanifolds of R^n and $\mathbb{R}^m$ (and therefore spaces $\mathbf{T}_p\mathscr{X}$ and $\mathbf{T}_q\mathscr{Y}$ can be naturally identified with these spaces), the mapping $df_p : \mathbf{T}_p\mathscr{X} \to \mathbf{T}_q\mathscr{Y}$ obviously coincides with the principal linear part $\mathbb{R}^n \to \mathbb{R}^m$ of f (i.e. is its differential in the sense of elementary analysis).

Let f: $\mathscr{X} \to \mathscr{Y}$ and g: $\mathscr{Y} \to \mathscr{Z}$ be smooth mappings. If $(U, h) = (U, x^1, \ldots, x^n)$, $(V, k) = (V, y^1, \ldots, y^m)$ and $(W, l) = (W, z^1, \ldots, z^s)$ are charts of manifolds $\mathscr{X}$, $\mathscr{Y}$, and $\mathscr{Z}$ such that $fU \subset V$ and $gV = W$ and if $\mathbf{y} = \mathbf{f}(\mathbf{x})$ and $\mathbf{z} = \mathbf{g}(\mathbf{y})$ are functions that give in these charts mappings f and g (we use a shorter notation introduced above), then the smooth mapping $g \circ f$: $\mathscr{X} \to \mathscr{Z}$ is obviously given in the charts (U, h) and (W, l) by the function $\mathbf{z} = \mathbf{g}(\mathbf{f}(\mathbf{x}))$. For any point $p \in U$ therefore by virtue of the rule on the differentiation of composite functions the equations

$$\left(\frac{\partial z^k}{\partial x^i}\right)_p = \left(\frac{\partial z^k}{\partial y^j}\right)_q \left(\frac{\partial y^j}{\partial x^i}\right)_p, \quad q = f(p),$$

hold which imply that the linear mapping $d(g \circ f)_p$ is a composition of linear mappings df_p and dg_q

$$d(g \circ f)_p = dg_q \circ df_p. \tag{23}$$

This formula is called the **chain rule**.

Of particular interest are two cases where $\mathscr{Y} = \mathbb{R}$ or $m = n$.

In the case where $\mathscr{Y}$ is the axis $\mathbb{R}$ (and hence the mapping f is a smooth function on $\mathscr{X}$) the differential df_p is usually called the *gradient* of the function f.

By virtue of identification (7) the gradient is a linear mapping $\mathbf{T}_p\mathscr{X} \to \mathbb{R}$, i.e. a covector of the space $\mathbf{T}_p\mathscr{X}$ (a vector of the conjugate space $\mathbf{T}_p^*\mathscr{X}$ which is generally called a *cotangent space* to a manifold $\mathscr{X}$ at a point p). By definition a covector df_p on any vector $A \in \mathbf{T}_p\mathscr{X}$ assumes a value

$$df_p(A) = \left(\frac{\partial f}{\partial x^i}\right)_p a^i. \tag{24}$$

This value is called the *derivative of a function f with respect to a vector A* and is denoted by Af. Thus

$$Af = df_p(A)$$

and

$$Af = \left(\frac{\partial f}{\partial x^i}\right) a^i \text{ if } A = (a^1, \ldots, a^n) \text{ in } (U, x^1, \ldots, x^n).$$

In particular, $\left(\frac{\partial}{\partial x^i}\right)_p f = \left(\frac{\partial f}{\partial x^i}\right)_p$ for any $i = 1, \ldots, n$, which explains the choice of notation for the vectors of basis (5).

Formula (24) implies that in the basis for the space $\mathbf{T}_p^*\mathscr{X}$ conjugate to basis (5) for the space $\mathbf{T}_p\mathscr{X}$ the covector df_p has the coordinates $\left(\frac{\partial f}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial f}{\partial x^n}\right)_p$. Therefore, first, this basis consists of covectors

$$dx_p^1, \ldots, dx_p^n$$

(clearly, the vector df_p is also defined for functions f given only in some neighbourhood of a point p) and, second,

$$df_p = \left(\frac{\partial f}{\partial x^1}\right)_p dx_p^1 + \ldots + \left(\frac{\partial f}{\partial x^n}\right)_p dx_p^n = \left(\frac{\partial f}{\partial x^i}\right)_p dx_p^i.$$

Remark 1. Note that a *gradient is a covector*. The notion, known from calculus, of the gradient of a smooth function on $\mathbb{R}^n$ as a vector (see Lecture 24 below) rests on an implicit identification of vectors and covectors through the standard Euclidean structure on $\mathbb{R}^n$.

Now let $n = m$.

Definition 3. A smooth mapping $f: \mathscr{X} \to \mathscr{Y}$ of manifolds of the same dimension is said to be *étale at a point* $p \in \mathscr{X}$ (or a *local diffeomorphism*) if on some neighbourhood U of that point it is a diffeomorphism of the neighbourhood onto a neighbourhood $V = fU$ of a point $q = f(p)$.

Any diffeomorphism $\mathscr{X} \to \mathscr{Y}$ is an étale mapping (at every point $p \in \mathscr{X}$). Conversely, according to the calculus theorem on the differentiability of the inverse function *every étale bijective mapping* $f: \mathscr{X} \to \mathscr{Y}$ *is a diffeomorphism*. Simple examples (see Remark 1 of Lecture 6) show that the condition that the mapping should be bijective is necessary here (i.e. it does not follow from the fact that the mapping is étale).

From formula (12) (applied to a diffeomorphism $f|_U: U \to V$ and to the inverse diffeomorphism $g: V \to U$) it immediately follows (since it is obviously the identity mapping that serves as the differential of the identity mapping) that the *differential* $df_p: \mathbf{T}_p\mathscr{X} \to \mathbf{T}_q\mathscr{Y}$ *of an arbitrary mapping* $f: \mathscr{X} \to \mathscr{Y}$ *étale at a point* p *is an isomorphism* (an invertible linear mapping). Conversely, if the *differential* $df_p: \mathbf{T}_p\mathscr{X} \to \mathbf{T}_q\mathscr{Y}$ *of a smooth mapping* $f: \mathscr{X} \to \mathscr{Y}$ *at a point* $p \in \mathscr{X}$ *is an isomorphism, then the mapping* f *is étale at the point* p. Indeed, the statement that the mapping df_p is an isomorphism implies (for given charts (U, h) and (V, k) with $p \in U$ and $fU \subset V$) that the determinant of matrix (9) is nonzero. Since this determinant is nothing but the Jacobian $D\varphi$ of a smooth mapping $\varphi = k \circ f \circ h^{-1}: h(U) \to f(V)$, it follows by virtue of the inverse mapping theorem (see Lecture 6) that on some neighbourhood W' of a point $h(p) \in h(U)$

in $h(U)$ the mapping φ is a diffeomorphism. The mapping f is therefore a diffeomorphism on the neighbourhood $U' = h^{-1}W'$ of a point p in $\mathscr{X}$, and hence is étale at p. $\square$

We have thus proved the following proposition known as the **étale mapping theorem.**

Proposition 1. *A smooth mapping $f: \mathscr{X} \to \mathscr{Y}$ is étale at a point $p \in \mathscr{X}$ if and only it its differential*

$$df_p: \mathbf{T}_p\mathscr{X} \to \mathbf{T}_q \mathscr{Y}, \quad q = f(p),$$

is an isomorphism. $\square$

If a mapping $f: \mathscr{X} \to \mathscr{Y}$ étale at a point p is a diffeomorphism on the support U of a chart (U, h), then a pair $(V, k) = (fU, h \circ (f|_U)^{-1}$ is obviously a chart in $\mathscr{Y}$. The mapping $\varphi: h(U) \to k(V)$ which conforms to the mapping f is the identity mapping, i.e. in the corresponding local coordinates $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$ the mapping f is given by formulas of the form

$$y^i = x^i, \quad i = 1, \ldots, n. \tag{25}$$

Thus *if a mapping $f: \mathscr{X} \to \mathscr{Y}$ is étale at a point $p \in \mathscr{X}$, then in manifolds $\mathscr{X}$ and $\mathscr{Y}$ there are local coordinates $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$ having the property that in these coordinates the mapping f is written by formulas* (25) (i.e. is a mapping which preserves the coordinates of vectors in another system of coordinates).

Proposition 1 is essentially just another version of the inverse mapping theorem which allows a fundamentally different embodiment.

Let $(U, h) = (U, x^1, \ldots, x^n)$ be an arbitrary chart of a smooth manifold $\mathscr{X}$ which is centred at a point $p_0 \in \mathscr{X}$ (i.e. such that $p_0 \in U$), and let

$$x^{i'} = x^{i'}(x^1, \ldots, x^n), \; i' = 1', \ldots, n', \tag{26}$$

be smooth functions given in some neighbourhood $U' \subset U$ of a point p_0.

Recall (see Lecture 6) that $x^{i'}$ in (26) has a double meaning: on the left it designates a function on U' and on the right a function on a neighbourhood $h(U')$ of a point $\mathbf{x}_0 = h(p_0)$ in $\mathbb{R}^n$. Accordingly the Jacobian

$$(27) \qquad \det\left\| \frac{\partial x^{i'}}{\partial x^i} \right\|, \quad \begin{matrix} i = 1, \dots, n, \\ i' = 1', \dots, n', \end{matrix}$$

of functions (26) can also be considered as either a function on U' or a function on $h(U')$. Its actual meaning must be understood from the context.

If functions (26) are local coordinates in a neighbourhood U', then at any point of that neighbourhood and, in particular, at a point p_0 Jacobian (27) is nonzero. The converse is as follows:

Proposition 2. *If Jacobian* (27) *is nonzero at a point* p_0, *then in some neighbourhood* $U'' \subset U'$ *of that point functions* (26) *are local coordinates.*

Proof. Functions (26) considered as functions on an open set $h(U')$ of $\mathbb{R}^n$ define some mapping φ of $h(U')$ into $\mathbb{R}^n$. If their Jacobian (27) is nonzero at a point $\mathbf{x}_0 = h(p_0)$ of $h(U')$, then by the inverse mapping theorem there is a neighbourhood V' of the point $\mathbf{x}_0$ (contained in the open set $h(U')$) on which the mapping φ is a diffeomorphism onto some open set of $\mathbb{R}^n$. Then the pair (U'', h''), where

$$U'' = h^{-1}V, \quad h'' = (\varphi\,|_{V'}) \circ (h\,|_{U''}),$$

is a chart on $\mathscr{X}$ connected with chart (U, h) by transition functions (26). Hence functions (26), but now considered as functions on U'', are local coordinates corresponding to a coordinate mapping h''. □

Proposition 2 is known as the **theorem on change of local coordinates.**

Now we return to arbitrary smooth mappings $f\colon \mathscr{X} \to \mathscr{Y}$, where $\dim \mathscr{X} = n$ and $\dim \mathscr{Y} = m$.

Definition 4. By the *rank* of a smooth mapping $f\colon \mathscr{X} \to \mathscr{Y}$ at a point $p \in \mathscr{X}$ is meant the rank r of a linear mapping $df_p\colon \mathbf{T}_p\mathscr{X} \to \mathbf{T}_q\mathscr{Y}$, $q = f(p)$, i.e. the rank of the Jacobian matrix (21) of mapping f at p.

It is clear that

$$0 \leqslant r \leqslant \min(n, m).$$

Since with a slight change in the elements of the matrix its rank can only increase, the rank of f at an arbitrary point of a sufficiently small neighbourhood of the point p is at least its rank at a point p_0. It may well be higher however.

Definition 5. A mapping $f: \mathscr{X} \to \mathscr{Y}$ is said to be *locally flat* at a point $p \in \mathscr{X}$ if there is a neighbourhood U of p on which the rank of f is constant (equal to a rank r at p).

Proposition 3. *If a mapping $f: \mathscr{X} \to \mathscr{Y}$ is locally flat at a point $p \in \mathscr{X}$, then in manifolds $\mathscr{X}$ and $\mathscr{Y}$ there exist charts $(U, x^1, \ldots, x^n)$ and $(V, y^1, \ldots, y^m)$ such that $p \in U$, $fU \subset V$, and f can be written in local coordinates $x^1, \ldots, x^n$ and $y^1, \ldots, y^m$ as follows:*

$$y^j = \begin{cases} x^j & \text{if } j = 1, \ldots, r, \\ 0 & \text{if } j = r+1, \ldots, m. \end{cases} \tag{28}$$

(Notice that coordinates $x^{r+1}, \ldots, x^n$ are not involved in formulas (28).)

Intuitively Proposition 3 states that near a point p a mapping locally flat at p is constructed as a projection $\mathbb{R}^n \to \mathbb{R}^r \subset \mathbb{R}^m$ of $\mathbb{R}^n$ along the last $n - r$ coordinate axes onto a coordinate subspace $\mathbb{R}^r$ of $\mathbb{R}^m$ consisting of points whose last $m - r$ coordinates are zero.

We prove Proposition 3 in the next lecture.

Lecture 13

Proof of the theorem on locally flat maps · Immersions and submersions · Submanifolds of a smooth manifold · A subspace tangent to a submanifold · Giving locally a submanifold · The uniqueness of the submanifold structure · The case of embedded submanifolds · The theorem on the inverse image of a regular value · Solutions of systems of equations · The group SL (n) *as a submanifold*

We prove Proposition 3 of the preceding lecture.

First let

$(U, h) = (U, x^1, \ldots, x^n)$ and $(V, k) = (V, y^1, \ldots, y^m)$ be arbitrary charts of manifolds $\mathscr{X}$ and $\mathscr{Y}$ having the property that $p \in U$ and $fU \subset V$, and let the mapping f be written in these charts as follows:

$$y^j = f^j(x^1, \ldots, x^n), \quad j = 1, \ldots, m.$$

Under the hypothesis a rectangular matrix consisting of numbers $\left(\frac{\partial f^j}{\partial x^i}\right)_p$ has a rank r. Indexing coordinates, if necessary, we may assume without loss of generality that this matrix has a nonzero minor

$$\det \left\| \left(\frac{\partial f^j}{\partial x^i}\right)_p \right\|, \quad i, j = 1, \ldots, r, \tag{1}$$

lying at the intersection of the first r rows and columns.

With this in mind, consider in a neighbourhood U of a point p functions $x^{1'}, \ldots, x^{n'}$ given by

$$x^{i'} = \begin{cases} f^i(x^1, \ldots, x^n) & \text{if } i = 1, \ldots, r, \\ x^i & \text{if } i = r+1, \ldots, n. \end{cases} \tag{2}$$

For these functions the Jacobian det $\left\| \frac{\partial x^{i'}}{\partial x^i} \right\|$ is obviously equal to determinant (1) and hence is other than zero. Consequently, according to Proposition 2 of Lecture 12 functions (2) are local coordinates in some neighbourhood U' of a point p.

Now we take a neighbourhood U' as U and coordinates $x^{1'}, \ldots, x^{n'}$ as $x^1, \ldots, x^n$, and thus obtain charts (U, h) and (V, k) in which the mapping f can be written as follows:

$$y^j = \begin{cases} x^j & \text{if } j = 1, \ldots, r, \\ f^j(x^1, \ldots, x^n) & \text{if } j = r+1, \ldots, n. \end{cases} \tag{3}$$

Consequently, the Jacobian matrix of f is in these charts of the form

$$\left\| \begin{matrix} 1 & \ldots & 0 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \ldots & 1 & 0 & \ldots & 0 \\ \frac{\partial f^{r+1}}{\partial x^1} & \ldots & \frac{\partial f^{r+1}}{\partial x^r} & \frac{\partial f^{r+1}}{\partial x^{r+1}} & \ldots & \frac{\partial f^{r+1}}{\partial x^n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f^m}{\partial x^1} & \ldots & \frac{\partial f^m}{\partial x^r} & \frac{\partial f^m}{\partial x^{r+1}} & \ldots & \frac{\partial f^m}{\partial x^n} \end{matrix} \right\|.$$

Since under the hypothesis the rank of this matrix is equal to r not only at a point p but also in some neighbourhood of that point (which may be assumed without loss of generality to coincide with U), all the derivatives

$$\begin{matrix} \frac{\partial f^{r+1}}{\partial x^{r+1}}, & \ldots, & \frac{\partial f^{r+1}}{\partial x^n} \\ \cdot & \cdot & \cdot \\ \frac{\partial f^m}{\partial x^{r+1}}, & \ldots, & \frac{\partial f^m}{\partial x^n} \end{matrix}$$

are identically zero in U. Since U may obviously be considered a connected neighbourhood, it follows that in U (i.e. more exactly in $h(U)$) the functions $f^{r+1}, \ldots, f^m$

are independent of $x^{r+1}, \ldots, x^n$, i.e. their values at any point $(x^1, \ldots, x^r, \ldots, x^n) \in h(U)$ can be written as

$$f^{r+1}(x^1, \ldots, x^r), \ldots, f^m(x^1, \ldots, x^r).$$

In other words, these functions may be considered to be functions on a set $\operatorname{pr} h(U) \subset \mathbb{R}^r$ consisting of points $(x^1, \ldots, x^r) \in \mathbb{R}^r$ such that $(x^1, \ldots, x^r, x^{r+1}, \ldots, x^n) \in h(U)$ at least with one choice of numbers $x^{r+1}, \ldots, x^n$ (geometrically a set $\operatorname{pr}(U) \subset \mathbb{R}^n$ is nothing but a *projection* of a set $h(U) \subset \mathbb{R}^n$ onto a subspace $\mathbb{R}^r \subset \mathbb{R}^n$).

On the other hand, it follows directly from formulas (3) that a similarly defined projection $\operatorname{pr} k(V) \subset \mathbb{R}^r$ of a set $k(V) \subset \mathbb{R}^m$ contains a set $\operatorname{pr} h(U)$. Therefore on some neighbourhood of a point $q = f(p)$ (namely, on the neighbourhood $k^{-1}((\operatorname{pr} h(U)) \times \mathbb{R}^{m-r})$) the functions

$$f^{r+1}(y^1, \ldots, y^r), \ldots, f^m(y^1, \ldots, y^r)$$

are defined. We put

$$y^{j'} = \begin{cases} y^j & \text{if } j = 1, \ldots, r, \\ y^j - f^j(y^1, \ldots, y^r) & \text{if } j = r+1, \ldots, m. \end{cases}$$

Since

$$\det \left\| \frac{\partial y^{j'}}{\partial y^j} \right\| = \begin{vmatrix} 1 \ldots 0 & 0 \ldots 0 \\ \ldots\ldots & \ldots\ldots \\ 0 \ldots 1 & 0 \ldots 0 \\ & 1 \ldots 0 \\ ? & \ldots\ldots \\ & 0 \ldots 1 \end{vmatrix} = 1,$$

where ? designates elements of no interest to us, functions $y^{1'}, \ldots, y^{n'}$ are, according to Proposition 2 of Lecture 12, local coordinates in some neighbourhood V' of a point q. The mapping f is given in charts $(U, x^1, \ldots, x^n)$ and $(V', y^{1'}, \ldots, y^{n'})$ by

$$y^{j'} = \begin{cases} x^j & \text{if } j = 1, \ldots, r, \\ 0 & \text{if } j = r+1, \ldots, m. \end{cases}$$

To complete the proof it remains to denote V' again by V and $y^{1'}, \ldots, y^{n'}$ by $y^1, \ldots, y^n$. □

Definition 1. A smooth mapping $f\colon \mathscr{X} \to \mathscr{Y}$ of a smooth n-dimensional manifold $\mathscr{X}$ into a smooth m-dimensional manifold $\mathscr{Y}$ is said to be an *immersion* at a point $p \in \mathscr{X}$ if its rank at that point is n (which is, of course, possible only for $n \leqslant m$), i.e. if the mapping

$$df_p\colon \mathbf{T}_p\mathscr{X} \to \mathbf{T}_q\mathscr{Y}, \quad q = f(p), \tag{4}$$

is a monomorphism.

Similarly, a mapping $f\colon \mathscr{X} \to \mathscr{Y}$ is said to be a *submersion* at a point $p \in \mathscr{X}$ if its rank at that point is m (and hence $n \geqslant m$), i.e. if mapping (4) is an epimorphism.

The mapping f is thus either an immersion or a submersion if its rank assumes the greatest possible value (for the given n and m). Therefore immersions and submersions are also called *mappings of maximal rank.*

According to Proposition 1 of Lecture 12 a mapping is both an immersion and a submersion (for $n = m$) if and only if it is an étale mapping.

A mapping $f\colon \mathscr{X} \to \mathscr{Y}$ which is an immersion (a submersion) at every point $p \in \mathscr{X}$ is called an *immersion* (a *submersion*).

It is clear that a submersion and an immersion, being mappings of maximal rank, are locally flat at p. According to Proposition 3 of Lecture 7 therefore *for any submersion $f\colon \mathscr{X} \to \mathscr{Y}$ at a point $p \in \mathscr{X}$ there are charts $(U, x^1, \ldots, x^n)$ and $(V, y^1, \ldots, y^m)$ of manifolds $\mathscr{X}$ and $\mathscr{Y}$ such that $p \in U$, $fU \subset V$, and the mapping f can be written in local coordinates $x^1, \ldots, x^n$ and $y^1, \ldots, y^n$ as follows*:

$$y^1 = x^1, \ldots, y^m = x^m. \tag{5}$$

Intuitively this implies that in the corresponding coordinates any submersion is locally represented by projection $\mathbb{R}^n \to \mathbb{R}^m$ carrying a point $(x^1, \ldots, x^m, \ldots, x^n) \in \mathbb{R}^n$ into a point $(x^1, \ldots, x^m) \in \mathbb{R}^m$.

For immersions we interchange the symbols and assume f to be a mapping $\mathscr{Y} \to \mathscr{X}$. Then *for any immersion $f\colon \mathscr{Y} \to \mathscr{X}$ at a point $q \in \mathscr{Y}$ there are charts $(V, y^1, \ldots, y^m)$ and $(U, x^1, \ldots, x^n)$ of manifolds $\mathscr{Y}$ and $\mathscr{X}$ such that $q \in V$, $fV \subset U$, and the mapping f can be written in local* *coordinates* $y^1, \ldots, y^m$ *and* $x^1, \ldots, x^n$ *by*

$$(6)\quad x^1 = y^1, \ldots, x^m = y^m, x^{n+1} = 0, \ldots, x^n = 0.$$

Intuitively this implies that in the corresponding coordinates any immersion can be locally represented by an embedding $\mathbb{R}^m \to \mathbb{R}^n$ sending a point $(y^1, \ldots, y^n) \in \mathbb{R}^m$ into a point $(y^1, \ldots, y^m, 0, \ldots, 0) \in \mathbb{R}^n$.

Definition 2. A smooth manifold $\mathscr{Y}$ is said to be a *submanifold* of a smooth manifold $\mathscr{X}$ if it is contained in $\mathscr{X}$ and the corresponding mapping of an embedding

$$(7)\qquad \iota: \mathscr{Y} \to \mathscr{X}, \quad \iota(p) = p,$$

is an immersion at any point $p \in \mathscr{Y}$ (and, in particular, a smooth mapping).

Since $\mathscr{Y} \subset \mathscr{X}$, on $\mathscr{Y}$ a topology $\mathbb{T}_{\mathscr{Y}/\mathscr{X}}$ induced by the topology $\mathbb{T}_{\mathscr{X}}$ of a manifold $\mathscr{X}$ is defined. This topology (see below) **differs** generally from the topology $\mathbb{T}_{\mathscr{Y}}$ of a manifold $\mathscr{Y}$. One can only say that $\mathbb{T}_{\mathscr{Y}/\mathscr{X}} \subset \mathbb{T}_{\mathscr{Y}}$ (this inclusion is equivalent to the continuity of mapping ι).

Definition 3. If ι is a homeomorphism onto its image (i.e. if $\mathbb{T}_{\mathscr{Y}/\mathscr{X}} = \mathbb{T}_{\mathscr{Y}}$), the manifold $\mathscr{Y}$ is said to be an *embedded submanifold.* Otherwise $\mathscr{Y}$ is said to be an *immersed submanifold.* (The latter term, however, is also used as a synonym of the term "submanifold" to emphasize that the submanifold $\mathscr{Y}$ is not embedded.)

Immersion (7), is of course, an injective mapping. Conversely let $f: \mathscr{Y} \to \mathscr{X}$ be an arbitrary immersion which is an injective mapping and let $f = \iota \circ f'$ be its decomposition into the product of a bijective $f': \mathscr{Y} \to f(\mathscr{Y})$ and an injective $\iota: f(\mathscr{Y}) \to \mathscr{X}$ mapping. Since f' is a bijective mapping, we may use it to carry the smoothness from $\mathscr{Y}$ onto $\mathscr{Y}' = f(\mathscr{Y})$. Then $\mathscr{Y}'$ will be a smooth manifold, f' a diffeomorphism, and $\iota = f \circ (f')^{-1}$ an immersion (as a composition of an immersion and a diffeomorphism), i.e. $\mathscr{Y}'$ will be a submanifold of a manifold $\mathscr{X}$.

Thus submanifolds of a manifold $\mathscr{X}$ are precisely the images in $\mathscr{X}$ of arbitrary immersions $\mathscr{Y} \to \mathscr{X}$ which are the injective mappings. The embedded submanifolds are therewith the images of the immersions which are moneomorphisms (homeomorphisms onto their own image). Examples of embedded submanifolds are simple regular arcs and elementary surfaces. Parametrizations are the corresponding immersions. The condition that parametrizations should be regular means that a parametrization is an immersion.

Since for any submanifold $\mathscr{Y} \subset \mathscr{X}$ embedding (7) is an immersion, for any point $p \in \mathscr{Y}$ the mapping

$$d\iota_p\colon \mathbf{T}_p \mathscr{Y} \to \mathbf{T}_p\mathscr{X}$$

is an isomorphism of a vector space $\mathbf{T}_p\mathscr{Y}$ onto a linear subspace $\operatorname{Im} d\iota_p = (d\iota_p)(\mathbf{T}_p\mathscr{Y})$ of a space $\mathbf{T}_p\mathscr{X}$. The subspace is called a *tangent space to a submanifold* $\mathscr{Y}$ *at a point* p. It is usually identified with $\mathbf{T}_p\mathscr{Y}$ (through an isomorphism $d\iota_p$).

Some peculiarities arise in the case where $\mathscr{X}$ is an affine space $\mathscr{A}$, and therefore for any point $p \in \mathscr{X}$ (and, in particular, for any point $p \in \mathscr{Y}$) the vector space $\mathbf{T}_p \mathscr{X}$ is identified with the associated vector space $\mathscr{V}$. In this case a subspace $\mathbf{T}_p\mathscr{Y}$ is usually identified with a linear submanifold $p + \mathbf{T}_p\mathscr{Y}$ of $\mathscr{A}$ (i.e. graphically, its vectors are laid off from a point p). Cf. Definition 3 of Lecture 3.

Being an immersion, mapping (7) can be written in local coordinates by formulas (6). This means that for any point p of a submanifold $\mathscr{Y}$ there is a chart $(U, x^1, \ldots, x^n)$ $p \in U$, of a manifold $\mathscr{X}$ such that, first, on some set $V \subset U \cap \mathscr{Y}$ (open in $\mathscr{Y}$) the restrictions

$$y^1 = x^1 |_V \,, \ldots, y^m = x^m |_V$$

of the first m coordinates $x^1, \ldots, x^m$ are local coordinates on V and, second, a point $q \in U$ is in V if and only if

$$x^{m+1}(q) = 0, \ldots, x^n(q) = 0. \tag{8}$$

Such coordinates $x^1, \ldots, x^n$ will be called *compatible* with a submanifold $\mathscr{Y}$.

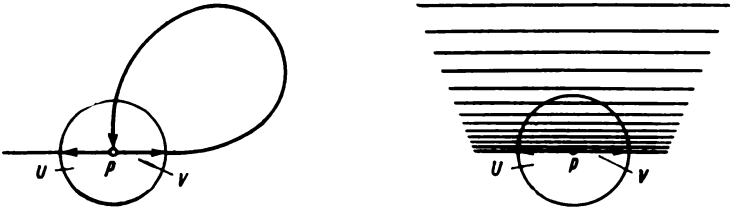
Coordinates $y^1, \ldots, y^m$ define in $\mathbf{T}_p\mathscr{Y}$ a basis

$$\left(\frac{\partial}{\partial y^1}\right)_p, \ldots, \left(\frac{\partial}{\partial y^m}\right)_p \tag{9}$$

and coordinates $x^1, \ldots, x^n$ in $\mathbf{T}_p\mathscr{X}$ a basis

$$\left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^n}\right)_p. \tag{10}$$

Since the embedding $\iota\colon \mathscr{Y} \to \mathscr{X}$ can be written in these coordinates by functions $y^1 = x^1, \ldots, y^m = x^m$, its differential $d\iota_p$ carries basis (9) into the first m vectors of basis (10). By virtue of $\mathbf{T}_p\mathscr{Y}$ being identified with $\operatorname{Im} d\iota_p$ it means that



Two typical examples where the equality $V = U \cap \mathscr{Y}$ cannot be satisfied

$$\left(\frac{\partial}{\partial y^1}\right)_p = \left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial y^m}\right)_p = \left(\frac{\partial}{\partial x^m}\right)_p,$$

i.e. *vectors* $\left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^m}\right)_p$ *make up a basis for a subspace* $\mathbf{T}_p\mathscr{Y}$ *of a space* $\mathbf{T}_p\mathscr{X}$.

If a submanifold $\mathscr{Y}$ is embedded in $\mathscr{X}$, then it may be assumed without loss of generality that

$$V = U \cap \mathscr{Y}. \tag{11}$$

Indeed, since V is open in $\mathscr{Y}$ and the topology in $\mathscr{Y}$ is induced by the topology of $\mathscr{X}$, there is an open set W in $\mathscr{X}$ such that $V = W \cap \mathscr{Y}$. Replacing U by $U \cap W$ we make, without loss of generality, equation (11) hold. □

Equation (11) implies that a point $q \in U$ is in $\mathscr{Y}$ if and only if equations (8) hold for it. In other words, (i.e. in a neighbourhood U) a manifold $\mathscr{Y}$ is given *locally* by $n - m$ equations

$$x^{m+1} = 0, \ldots, x^n = 0. \tag{12}$$

Generally it is impossible to make equation (11) hold for an unembedded (immersed) submanifold $\mathscr{Y}$, and the relation between V and $U \cap \mathscr{Y}$ becomes more complicated.

Since functions $x^{m+1}, \ldots, x^n$ are continuous, and hence conditions (8) single out in U a closed subset, the set V is closed in U (with respect to the topology induced in U by the topology $\mathbb{T}_{\mathscr{Y}/\mathscr{X}}$). Consequently, V is also closed in $U \cap \mathscr{Y}$ (with respect to the topology induced by the topology $\mathbb{T}_{\mathscr{Y}/\mathscr{X}}$ and hence with respect to the topology induced by the topology $\mathbb{T}_{\mathscr{Y}}$). On the other hand, under the hypothesis the set V is open in $\mathscr{Y}$ and hence in $U \cap \mathscr{Y}$ (with respect to the topology induced by the topology $\mathbb{T}_{\mathscr{Y}}$). The set V is thus open-closed in $U \cap \mathscr{Y}$ (with respect to the topology induced by the topology $\mathbb{T}_{\mathscr{Y}}$). Consequently, *if a coordinate neighbourhood V is connected, then it is a component of the set $U \cap \mathscr{Y}$ that contains a point p.*

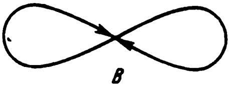
[Recall that any point of a smooth manifold has a fundamental system of connected coordinate neighbourhoods homeomorphic, for example, to open balls of a Euclidean space of the corresponding dimension.]

Now suppose that two smoothnesses are given on a subset $\mathscr{Y}$ of a smooth manifold $\mathscr{X}$ with respect to which it is a submanifold of $\mathscr{X}$ and which define on $\mathscr{Y}$ *the same topology*. Then charts of the form $(V, y^1, \ldots, y^m)$ with the properties described above are the same for both smoothnesses (since functions $y^1, \ldots, y^m$ are characterized as restrictions of local coordinates $x^1, \ldots, x^m$ and sets V as components of sets $U \cap \mathscr{Y}$). On the other hand, in each smoothness these charts obviously make up a subatlas of a maximal atlas. Therefore both smoothnesses coincide.

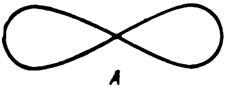
Thus *for a given topology on a subset $\mathscr{Y}$ of a smooth manifold $\mathscr{X}$ there can exist at most one smooth manifold structure on $\mathscr{Y}$ with respect to which $\mathscr{Y}$ is a submanifold of the manifold $\mathscr{X}$.*

Varying the topology we can obtain on $\mathscr{Y}$ many different submanifold structures. For example, any subset $\mathscr{Y} \subset \mathscr{X}$ can be provided with a discrete topology and thus turned into a zero-dimensional submanifold.

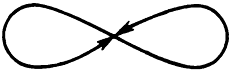
A more interesting example can be obtained by considering a set in a plane depicted in Fig. A (a "figure eight"). The set cannot be a submanifobd of the plane in the induced topology because of the singular point in the centre. But in a weaker topology conventionally represented in Fig. B the same set is an immersed manifold diffeomorphic to the straight line $\mathbb{R}$. It is an immersed submanifold diffeomorphic to the straight line $\mathbb{R}$ in another topology conventionally represented in Fig. C. The corresponding embeddings are nonequivalent regular curves without double points having the same support.



B



A



C

In connection with the topology on $\mathscr{Y}$ induced by the topology of a manifold $\mathscr{X}$ we see, in particular, that *on a subset* $\mathscr{Y}$ *of a smooth manifold* $\mathscr{X}$ *there can be at most one smooth manifold structure with respect to which* $\mathscr{Y}$ *is an embedded submanifold.* We can therefore say that the subset $\mathscr{Y}$ is an embedded **submanifold.**

To determine whether a given subset $\mathscr{Y}$ is an embedded submanifold one should

1° Consider all possible pairs of the form $(V, y^1, \ldots, y^m)$, where V is the intersection $U \cap \mathscr{Y}$ of $\mathscr{Y}$ and the support U of some chart $(U, x^1, \ldots, x^n)$ of a manifold $\mathscr{X}$ and $y^1, \ldots, y^m$ are the restrictions $x^1|_V, \ldots, x^m|_V$ on V of local coordinates $x^1, \ldots, x^m$.

2° Choose pairs $(V, y^1, \ldots, y^m)$ from among these pairs which are charts on $\mathscr{Y}$, i.e. such that the functions $y^1, \ldots, y^m$ give a bijective mapping of a set $V = U \cap \mathscr{Y}$ onto some open set $V \subset \mathbb{R}^m$.

Therefore, *if the chosen charts can make up an atlas on* $\mathcal{Y}$, *then* $\mathcal{Y}$ *is an embedded submanifold.* Indeed, it is obvious that this atlas gives on $\mathcal{Y}$ a smoothness with respect to which $\mathcal{Y}$ is a submanifold and which is compatible with the induced topology. [Notice that in general the maximal atlas of a manifold $\mathcal{Y}$ can contain only some of the chosen charts.] □

We apply this general statement to a set $\mathcal{Y} = f^{-1}(q_0)$ which is a full inverse image of some point $q_0 \in \mathfrak{Z}$ under a smooth mapping $f\colon \mathcal{X} \to \mathfrak{Z}$.

Definition 4. A point $q_0 \in \mathfrak{Z}$ is said to be a *regular value* of a mapping $f\colon \mathcal{X} \to \mathfrak{Z}$ if f is a submersion at each point $p \in \mathcal{Y}$.

Proposition 1 (theorem on the inverse image of a regular value). *The inverse image* $\mathcal{Y} = f^{-1}(q_0)$ *of an arbitrary regular value* q_0 *(when* $\mathcal{Y}$ *is not empty) is an embedded submanifold of a manifold* $\mathcal{X}$. *The dimension of the submanifold is* $n - r$, *where* $r = \dim \mathfrak{Z}$ *(and* $n = \dim \mathcal{X}$*).*

Proof. For any point $p \in \mathcal{Y}$ in a manifold $\mathcal{X}$ there is a chart $(U, x^1, \ldots, x^n) = (U, h)$ and in $\mathfrak{Z}$ there is a chart $(W, z^1, \ldots, z^r)$ such that $p \in U$, $fU \subset W$, and in these charts a mapping can be written as follows

$$z^1 = x^{m+1}, \ldots, z^r = x^n, \tag{13}$$

where $m = n - r$. We may assume without loss of generality that at a point q_0 all the coordinates $z^1, \ldots, z^r$ are zero, and hence a point $u \in U$ is in the set $\mathcal{Y}$ (i.e. more exactly in the intersection $V = U \cap \mathcal{Y}$) if and only if

$$x^{m+1}(u) = 0, \ldots, x^n(u) = 0.$$

It follows that if a mapping $k\colon V \to \mathbb{R}^m$ given by

$$k(u) = (y^1(u), \ldots, y^m(u)), \quad u \in U,$$

where

$$y^1 = x^1|_V, \ldots, y^m = x^m|_V,$$

is to be considered due to the natural embedding

$$\mathbb{R}^m \to \mathbb{R}^n, \ (x^1, \ldots, x^m) \mapsto (x^1, \ldots, x^m, 0, \ldots, 0)$$

as a mapping in $\mathbb{R}^n$, then it is nothing but a restriction $h|_V$ on V of a mapping $h\colon U \to h(U)$. The set $k(V)$ being an intersection $\mathbb{R}^m \cap h(U)$ is open in $\mathbb{R}^m$, and the mapping $k\colon V \to k(V)$ is bijective. In other words, the pair $(V, k) = (V, y^1, \dots, y^m)$ is a chart in $\mathcal{Y}$.

Now let $(U', x^{1'}, \dots, x^{n'})$ be another chart of a manifold $\mathcal{X}$ having similar properties with respect to a mapping f and let $(V', y^{1'}, \dots, y^{m'})$ be the corresponding chart in $\mathcal{Y}$. Since the coordinates $x^1, \dots, x^n$ and $x^{1'}, \dots, x^{n'}$ are connected in the intersection $U \cap U'$ by formulas of the form

$$x^{i'} = x^{i'}(x^1, \dots, x^n), \quad i' = 1', \dots, n,$$

where

$$\det\left\|\frac{\partial x^{i'}}{\partial x^i}\right\| \neq 0 \text{ everywhere on } U \cap U', \tag{14}$$

the coordinates $y', \dots, y^m$ and $y^{1'}, \dots, y^{m'}$ are connected in the intersection $V \cap V' = U \cap U' \cap \mathcal{Y}$ by

$$y^{j'} = x^{j'}(y^1, \dots, y^m, 0, \dots, 0), \; j' = 1, \dots, m. \tag{15}$$

Besides, since in U' the set $\mathcal{Y}$ is defined by the equations

$$x^{m+1'} = 0, \dots, x^{n'} = 0,$$

in $V \cap V' \neq \varnothing$ there must also hold the relations

$$x^{k'}(y^1, \dots, y^m, 0, \dots, 0) = 0, \quad k' = m + 1, \dots, n.$$

Hence it follows that for partial derivatives $\dfrac{\partial x^{i'}}{\partial x^l}$, $l = 1, \dots, m$, the following equations

$$\frac{\partial x^{i'}}{\partial x^l} = \begin{cases} \dfrac{\partial y^{i'}}{\partial y^l} & \text{if } i' = 1, \dots, m, \\ 0 & \text{if } i' = m+1, \dots, n, \end{cases}$$

hold at each point of $V \cap V'$, and hence Jacobian (14) is divided by a Jacobian

$$\det\left\|\frac{\partial y^{i'}}{\partial y^l}\right\|, \quad \begin{matrix} i' = 1', \dots, m', \\ l = 1, \dots, m. \end{matrix} \tag{16}$$

Consequently, Jacobian (16) is nonzero, and hence formulas (15) specify diffeomorphism of the corresponding sets.

This proves that any two charts of the form $(V, y^1, \dots, y^m)$ on $\mathcal{Y}$ are compatible. Since they obviously cover $\mathcal{Y}$, it follows that these charts make up an atlas, and hence $\mathscr{Y}$ is an embedded manifold. □

It follows directly from formulas (13) giving a mapping f in local coordinates that the differential

$$df_p\colon \mathbf{T}_p\mathscr{X} \to \mathbf{T}_{q_0}\mathfrak{Z}$$

of that mapping at a point $p \in \mathscr{Y}$ acts on the vectors of the basis for a space $\mathbf{T}_p\mathscr{X}$ as follows:

$$df_p\left(\frac{\partial}{\partial x^i}\right)_p = \begin{cases} 0 & \text{if } i = 1, \ldots, m, \\ \left(\frac{\partial}{\partial z^{i-m}}\right)_{q_0} & \text{if } i = m+1, \ldots, n. \end{cases}$$

Hence *vectors*

$$\left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^m}\right)_p \tag{17}$$

generate the kernel Ker df_p *of a mapping* df_p, *and vectors*

$$\left(\frac{\partial}{\partial x^{m+1}}\right)_p, \ldots, \left(\frac{\partial}{\partial x^n}\right)_p \tag{18}$$

generate a subspace mapped isomorphically onto a tangent space $\mathbf{T}_{q_0}\mathfrak{Z}$ *of a manifold* $\mathfrak{Z}$.

Thus

$$\mathbf{T}_p\mathscr{X} = \operatorname{Ker} df_p \oplus \hat{\mathbf{T}}_{q_0}\mathfrak{Z},$$

where $\hat{\mathbf{T}}_{q_0}\mathfrak{Z}$ is a subspace spanned by vectors (18).

On the other hand, since $y^1 = x^1|_V, \ldots, y^m = x^m|_V$, vectors (17) generate a subspace $\mathbf{T}_p\mathscr{Y}$ of a space $\mathbf{T}_p\mathscr{X}$. Therefore *for any point* $p \in \mathscr{Y}$ *the kernel* Ker df_p *of the mapping* df_p *coincides with the tangent space* $\mathbf{T}_p\mathscr{Y}$ *of a submanifold* $\mathscr{Y}$:

$$\mathbf{T}_p\mathscr{Y} = \operatorname{Ker} df_p, \tag{19}$$

and hence

$$\mathbf{T}_p\mathscr{X} = \mathbf{T}_p\mathscr{Y} \oplus \hat{\mathbf{T}}_{q_0}\mathfrak{Z}. \tag{20}$$

Note that this decomposition holds for any point $p \in \mathscr{Y}$.

Remark 1. It should be kept in mind that by no means any embedded submanifold $\mathscr{Y} \subset \mathscr{X}$ is the inverse image of a regular value under some mapping $f\colon \mathcal{Y} \to \mathcal{Z}$ (see Problem 1 below).

According to formula (20), for an embedded submanifold $\mathcal{Y} \subset \mathcal{X}$ to be the inverse image of a regular value it is necessary (and as can be shown sufficient) that for any point $p \in \mathcal{Y}$ there should hold a decomposition of the form

$$\mathbf{T}_p\mathcal{X} = \mathbf{T}_p\mathcal{Y} \oplus N_p,$$

where N_p is a subspace for which *for all* p an isomorphic mapping $N_p \to \mathbb{R}^{n-m}$ dependent smoothly (in the obvious sense) on a point p is given.

Let $f^1, \ldots, f^r$ be smooth functions on a smooth manifold $\mathcal{X}$, and $a^1, \ldots, a^r$ be real numbers. A point $p \in \mathcal{X}$ is said to be a *solution* of a system of equations

$$(21) \qquad f^1 = a^1, \ldots, f^r = a^r$$

if $f^i(p) = a^i$ for any $i = 1, \ldots, r$.

Let $\mathcal{Y}$ be the set of all solutions of system (21).

Equations (21) are said to be *functionally independent* if for any point $p \in \mathcal{Y}$ covectors

$$df^1_p, \ldots, df^r_p \in \mathbf{T}^*_p\mathcal{X}$$

are linearly independent, and hence equations

$$(22) \qquad df^1_p = 0, \ldots, df^r_p = 0$$

define in $\mathbf{T}_p\mathcal{X}$ some linear subspace of dimension $m = n - r$.

Proposition 2. *For every system* (21) *of functionally independent equations the set $\mathcal{Y}$ of its solutions is an embedded submanifold of a manifold $\mathcal{X}$ of dimension $m = n - r$.*

For every point $p \in \mathcal{Y}$ the subspace $\mathbf{T}_p\mathcal{Y}$ is a subspace of solutions of a system of linear equations (22).

Proof. Functions $f^1, \ldots, f^r$ define (by formula $f(p) = (f^1(p), \ldots, f^r(p))$ a smooth mapping $f\colon \mathcal{X} \to \mathbb{R}^r$ for which the set $\mathcal{Y}$ is the inverse image $f^{-1}(\mathbf{a})$ of a point $\mathbf{a} = (a^1, \ldots, a^r) \in \mathbb{R}^r$. The functional independence of equations (21) obviously implies that the point $\mathbf{a}$ is a regular value of a mapping f. The first statement of Proposition 2 is therefore just a restatement of Proposition 1 for the case where $\mathcal{Z} = \mathbb{R}^r$.

Similarly, since the space of solutions of equations (22) is nothing but the kernel of a mapping df_p, the second statement is a restatement of equation (19). □

Corollary 1. *Let f be a smooth function on a smooth manifold $\mathscr{X}$, and let $\mathscr{Y}$ be the set of all points $p \in \mathscr{X}$ for which $f(p) = a$, where $a \in \mathbb{R}$ is a fixed number. If $df_p \neq 0$ for every point $p \in \mathscr{Y}$, then $\mathscr{Y}$ is an embedded submanifold of dimension $n - 1$ whose tangent space at an arbitrary point $p \in \mathscr{Y}$ is a hyperplane $df_p = 0$.* □

For the function $f = f(x, y)$ on $\mathbb{R}^2$ the condition $df_p \neq 0$ implies that either $\frac{\partial f}{\partial x} \neq 0$ or $\frac{\partial f}{\partial y} \neq 0$. The condition that a subset in a plane with equation $f(x, y) = 0$ should be an embedded one-dimensional manifold implies precisely that the subset has no singular points (see Lecture 1).

Remark 2. As shown by the Whitney theorem in Lecture 1, it is practically impossible to get rid of the condition $df_p \neq 0$ in Corollary 1. Nevertheless this condition is by no means necessary for the set $\mathscr{Y}$ to be an embedded (and not necessarily $(n - 1)$-dimensional) submanifold. In $\mathbb{R}^3$, for example, the equation

$$x^2 + y^2 = 0$$

defines a straight line (an embedded submanifold), but at all points of the straight line the differential df_p of the function $f = x^2 + y^2$ is identically zero.

Moreover, according to Remark 1 it is by no means any submanifold that can be given by a system of functionally independent equations.

Problem 1. Show that on the projective plane $\mathbb{R}P^2$ there is no smooth function f whose set of zeros would be the projective straight line $\mathbb{R}P^1 \subset \mathbb{R}P^2$.

Nevertheless in practice the fact that one subset or another of a smooth manifold is an embedded submanifold would as a rule be established using Proposition 2 (or its Corollary 1).

Example 1. The determinant det A of a square matrix $A = \| a_{ij} \|$ is a polynomial in a_{ij} of degree 1 in each variable. It can easily be seen that

$$(23) \qquad \frac{\partial \det A}{\partial a_{ij}} = A_{ij},$$

where A_{ij} is the cofactor of an element a_{ij}. (It suffices to notice that by the determinant expansion formula by column elements $\det A = A_{ij}a_{ij} + A'_{ij}$, where A'_{ij} is independent of a_{ij}). Since for $\det A = 1$ it is required that $A_{ij} \neq 0$ for at least one element a_{ij}, it follows that the conditions of Corollary 1 to Proposition 2 hold for the equation

$$\det A = 1$$

(on the manifold GL (n)). Consequently, the *group* SL (n) *is an embedded submanifold of dimension* $n^2 - 1$ *of* GL (n).

Problem 2. Show that the smoothness on the group SL (n) as a submanifold of the manifold GL (n) coincides with the smoothness of the group as a matrix Lie group (see Lecture 5).

Since the group GL (n) is an open submanifold of the vector space $\mathrm{Mat}_n(\mathbb{R})$, for any matrix $A \in \mathrm{GL}(n)$ the tangent space $\mathbf{T}_A\mathrm{GL}(n)$ to GL (n) in A can be naturally identified with $\mathrm{Mat}_n(\mathbb{R})$. Corresponding to the basis vector $\left(\frac{\partial}{\partial a_{ij}}\right)_A$ of $\mathbf{T}_A\mathrm{GL}(n)$ is the matrix unit $E_{ij} \in \mathrm{Mat}_n(\mathbb{R})$, and hence corresponding to the covector $(da_{ij})_A$ is the covector $\mathrm{Mat}_n(\mathbb{R}) \to \mathbb{R}$ which associates a matrix $B = \| b_{ij} \|$ with its element b_{ij}.

It follows by virtue of (23) that the differential df_A of the function $f\colon A \mapsto \det A$ considered as a covector of the space $\mathrm{Mat}_n(\mathbb{R})$ acts by the formula

$$df_A(B) = \sum_{i,j=1}^{n} A_{ij}b_{ij}, \quad B = \| b_{ij} \| \in \mathrm{Mat}_n(\mathbb{R}),$$

where A_{ij} are the cofactors of the elements of A. In particular, for $A = E$ (when $A_{ij} = \delta_{ij}$) we get

$$df_E(B) = \sum_{i=1}^{n} b_{ii} = \mathrm{Tr}\, B,$$

and hence $\mathrm{Ker}\, df_E = \mathfrak{sl}(n)$.

Since Ker $df_E = \mathbf{T}_E\mathrm{SL}(n)$ this proves that the *tangent space to the subgroup* SL (n) *at a point E is its Lie algebra* $\mathfrak{sl}(n)$.

Remark 3. In the next semester* we show (see Lecture IV.13) that any matrix Lie group $\mathcal{G}$ is an embedded submanifold of the manifold GL (n) the tangent space to which at a point E is the Lie algebra $\mathfrak{g}$ of the group.

---

* M. M. Postnikov. *Lectures in Geometry*: *Semester* IV. *Differential Geometry*. Nauka Publishers, Moscow, 1988.

Lecture 14

The embedding theorem · Compact sets revisited · Urysohn functions · Proof of the embedding theorem · Manifolds satisfying the second axiom of countability · Scattered and meager sets · Null sets

---

The simplest, and the most obvious, class of manifolds are manifolds which, for some $N > 0$, are diffeomorphic to embedded submanifolds of the space $\mathbb{R}^N$ or, to simplify the definitions, manifolds that can be $\mathbb{R}^N$-*embedded.*

Clearly any subspace of a Hausdorff topological space satisfying the second axiom of countability is Hausdorff and satisfies the axiom. *Any* $\mathbb{R}^N$-*embedded manifold is therefore Hausdorff and satisfies the second axiom of countability.* It turns out that this necessary condition is also sufficient:

Theorem 1. *For any smooth Hausdorff manifold $\mathcal{X}$ satisfying the second axiom of countability there is an N such that the manifold $\mathcal{X}$ can be* $\mathbb{R}^N$-*embedded.*

What can we say about N?

Proposition 1. *If a smooth manifold of dimension n is* $\mathbb{R}^N$-*embedded, where $N > 2n + 1$, then it is also* $\mathbb{R}^{N-1}$-*embedded.*

Corollary (Whitney theorem). *Any smooth Hausdorff manifold $\mathcal{X}$ of dimension n satisfying the second axiom of countability can be* $\mathbb{R}^{2n+1}$-*embedded.*

We prove Theorem 1 only under the assumption that the *manifold $\mathcal{X}$ is compact.* [Notice that a compact manifold satisfies (prove it!) the second axiom of countability.] The general case requires additional technicalities for which we have no time.

Remark 1. Theorem 1 is true for manifolds of an arbitrary C^r class, $r \geqslant 1$, but we shall not prove it for C^1-manifolds and for (real analytic) C^ω- manifolds which require quite different, much more complicated considerations.

We begin with some simple remarks about compact sets.

It is easy to see that the *continuous image of a compact set is compact*, i.e. for any continuous mapping $f\colon \mathscr{X} \to \mathscr{Y}$ and any compact subset $C \subset \mathscr{X}$ the set $fC \subset \mathscr{Y}$ is compact. Indeed, if open sets U_α cover fC, then open sets $f^{-1}U_\alpha$ cover C, and if the sets $f^{-1}U_{\alpha_1}, \ldots, f^{-1}U_{\alpha_n}$ cover C, then the sets $U_{\alpha_1}, \ldots, U_{\alpha_n}$ cover fC. □

Generally, a compact subset may be nonclosed, but if the space is Hausdorff, then this is impossible, i.e. *in a Hausdorff space $\mathscr{X}$ any compact subset C is closed.* Indeed, since $\mathscr{X}$ is a Hausdorff space, for any point $c \in C$ and any point $p \in \mathscr{X} \setminus C$ there are neighbourhoods $U_p(c)$ and $V_c(p)$ for points c and p, respectively, such that $U_p(c) \cap V_c(p) = \varnothing$. Since for every point $p \in \mathscr{X} \setminus C$ all neighbourhoods of the form $U_p(c)$, $c \in C$, cover the set C, in view of the set being compact there is a finite system of points $c_1, \ldots, c_k \in C$ such that

$$C \subset U_p(c_1) \cup \ldots \cup U_p(c_k).$$

But then the intersection

$$V = V_{c_1}(p) \cap \ldots \cap V_{c_k}(p) \tag{1}$$

is a neighbourhood of p that does not intersect with C, i.e. contained in $\mathscr{X} \setminus C$. This proves that any point p of the set $\mathscr{X} \setminus C$ is its interior point, i.e. $\mathscr{X} \setminus C$ is an open set. Consequently, the set C is closed. □

It follows that *if a space $\mathscr{X}$ is compact and a space $\mathscr{Y}$ is a Hausdorff space, then every continuous mapping $f\colon \mathscr{X} \to \mathscr{Y}$ is closed* (for any closed set $C \subset \mathscr{Y}$ the set $fC \subset \mathscr{Y}$ is closed). Indeed, if the set C is closed, then it is compact since the space $\mathscr{X}$ is compact. The set fC is therefore compact, and hence closed since the space $\mathscr{Y}$ is Hausdorff. □

Of particular importance is the case where f is a bijective mapping. Since in this case the fact that the mapping f is closed implies that the inverse mapping f^{-1}: $\mathcal{Y} \to \mathcal{X}$ is continuous, we see that a *continuous bijective mapping of a compact space onto a Hausdorff space is a homeomorphism*.

If, however, a continuous mapping f: $\mathcal{X} \to \mathcal{Y}$ of a compact space $\mathcal{X}$ into a Hausdorff space $\mathcal{Y}$ is only injective, then it is a moneomorphism (a homeomorphism onto its image $f(\mathcal{Y})$ which in this case is closed in $\mathcal{Y}$).

Therefore, *every compact submanifold of a Hausdorff manifold is an embedded submanifold.*

Now let $\mathcal{X}$ be an arbitrary smooth manifold, and W and V be its open subsets such that

$$\overline{W} \subset V. \tag{2}$$

Definition 1. By *Urysohn's function* of a pair (V, W) we mean a smooth function φ: $\mathcal{X} \to \mathbb{R}$ such that

(a) $0 \leqslant \varphi(p) \leqslant 1$ for any point $p \in \mathcal{X}$,

(b) $\varphi(p) = 1$ if and only if $p \in \overline{W}$,

(c) if $p \in \mathcal{X} \setminus V$, then $\varphi(p) = 0$.

It can be shown that if the manifold $\mathcal{X}$ is Hausdorff and satisfies the second axiom of countability, then *Urysohn's function exists for any pair (V, W) of open sets that satisfy function* (2). However, we do not need this statement and shall restrict ourselves to a proof of the following weaker result.

Proposition 2. *If a smooth manifold $\mathcal{X}$ is Hausdorff, then for any of its points p_0 and any neighbourhood U of that point there are open sets W and V such that*

$$p_0 \in W, \quad \overline{W} \subset V, \quad \overline{V} \subset U$$

and for a pair (V, W) there is a Urysohn function.

Proof. It is clear that without loss of generality a neighbourhood U may be assumed to be a coordinate neighbourhood of a point p_0. This means by definition that there is a diffeomorphism h of the set U onto some open set $h(U) \subset \mathbb{R}^n$. Let $\mathbf{x}_0 = h(p_0)$. Since $\mathbf{x}_0 \in h(U)$ and the set $h(U)$ is open, there exists an $r > 0$ such that every point $\mathbf{x} \in \mathbb{R}^n$ for which $|\mathbf{x} - \mathbf{x}_0| < r$ (i.e. each point of a ball $\mathbf{B}_r^n(\mathbf{x}_0)$) is in $h(U)$.

Replacing the diffeomorphism h by its composition with some similarity transformation of $\mathbb{R}^n$, we may therefore assume without loss of generality that $\mathbf{x}_0 = 0$ and $r = 3$, i.e. that $\mathring{\mathbf{B}}_3^n \subset h(U)$, where (see Lecture 1) $\mathring{\mathbf{B}}_3^n$ is an open ball of $\mathbb{R}^n$ of radius 3 with centre at the point 0.

For any $r < 3$ in a manifold $\mathscr{X}$ a set $h^{-1}(\mathbf{B}_r^n)$ is then defined. Since the ball $\mathbf{B}_r^n$ is compact, and the mapping h^{-1} is a homeomorphism, this set is also compact and hence closed, since under the hypothesis the manifold $\mathscr{X}$ is Hausdorff.

On the other hand, the set $h^{-1}(\mathring{\mathbf{B}}_r^n)$ is open (in U and hence in $\mathscr{X}$ too) and its closure $\overline{h^{-1}(\mathring{\mathbf{B}}_r^n)}$ obviously contains the set $h^{-1}(\mathbf{B}_r^n)$. Therefore

$$\overline{h^{-1}(\mathring{\mathbf{B}}_r^n)} = h^{-1}(\mathbf{B}_r^n).$$

Putting

$$W = h^{-1}(\mathring{\mathbf{B}}_1^n) \text{ and } V = h^{-1}(\mathring{\mathbf{B}}_2^n)$$

we immediately get

$$\overline{W} \subset V \text{ and } \overline{V} \subset U,$$

with $p_0 \in W$ and

$$h(\overline{W}) = \mathbf{B}_1^n \text{ and } h(\overline{V}) = \mathbf{B}_2^n.$$

Now recall that according to Corollary 2 to Lemma 1 of Lecture 1 there is a smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$ on $\mathbb{R}^n$ such that $0 \leqslant f \leqslant 1$ on $\mathbb{R}^n$ and

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if and only if } \mathbf{x} \in \mathbf{B}_1^n, \\ 0 & \text{if and only if } \mathbf{x} \notin \mathring{\mathbf{B}}_2^n. \end{cases}$$

With all this in mind we define the function $\varphi : \mathscr{X} \to \mathbb{R}$ as follows:

$$\varphi(p) = \begin{cases} f(h(p)) & \text{if } p \in U, \\ 0 & \text{if } p \notin U. \end{cases} \tag{3}$$

Clearly, this function has properties (a), (b), and (c) of Definition 1. It is only necessary therefore to prove that function (3) is smooth.

By construction function (3) is smooth on U and on $\mathscr{X}\setminus\overline{V}$ (it is zero on the latter set). The open sets U and $\mathscr{X}\setminus\overline{V}$ cover $\mathscr{X}$. On the other hand, it is clear that if some function $\varphi : \mathscr{X} \to \mathbb{R}$ is smooth on any element U_α of an open covering $\{U_\alpha\}$ of the manifold $\mathscr{X}$, then it is smooth on the whole $\mathscr{X}$. In particular, function (3) is therefore smooth on $\mathscr{X}$. □

Now we can prove Theorem 1.

Proof of Theorem 1 (for a compact $\mathscr{X}$). Choosing for every point $p_0 \in \mathscr{X}$ a coordinate neighbourhood U and applying Proposition 2, consider a set W specified by this proposition. Constructed for all points $p_0 \in \mathscr{X}$, these sets make up an open covering of a manifold $\mathscr{X}$. Since under the hypothesis $\mathscr{X}$ is compact, we can therefore choose a finite subcovering in that covering. Denote the sets W that make up the subcovering by $W_1, \ldots, W_m$ and consider the corresponding sets

$$V_1, \ldots, V_m, \quad U_1, \ldots, U_m,$$

coordinate mappings

$$h_1\colon U_1 \to \mathbb{R}^n, \ldots, h_m\colon U_m \to \mathbb{R}^n$$

and functions

$$\varphi_1\colon \mathscr{X} \to \mathbb{R}, \ldots, \varphi_m\colon \mathscr{X} \to \mathbb{R}$$

of Proposition 2.

These sets, mappings, and functions have the following properties:

1° For any $i = 1, \ldots, m$ there exist embeddings

$$\overline{W}_i \subset V_i, \quad \overline{V}_i \subset U_i.$$

2° The families

$$\{(W_1, h_1), \ldots, (W_m, h_m)\},$$
$$\{(V_1, h_1), \ldots, (V_m, h_m)\},$$
$$\{(U_1, h_1), \ldots, (U_m, h_m)\}$$

are the atlases of a manifold $\mathscr{X}$. (In the first two cases, of course, the corresponding restrictions of the mappings $h_1, \ldots, h_m$ are meant.)

3° Every function φ_i, $i = 1, \ldots, m$, is the Urysohn function of a pair (V_i, W_i).

Writing vectors of $\mathbb{R}^{n+1}$ as pairs $(\mathbf{x}, x)$, where $\mathbf{x} \in \mathbb{R}^n$ and $x \in \mathbb{R}$, we define mappings

$$f_i\colon \mathscr{X} \to \mathbb{R}^{n+1}, \quad i = 1, \ldots, m,$$

as follows:

$$f_i(p) = \begin{cases} (\varphi_i(p)\,h_i(p), & \varphi_i(p)) \quad \text{if} \quad p \in U_i, \\ (0, 0) & \qquad\qquad \text{if} \quad p \notin U_i. \end{cases}$$

It is clear that the mapping f_i is smooth on U_i and zero outside $\overline{V}_i$. It is therefore smooth on the entire manifold $\mathscr{X}$. Besides, $f_i(p) = (h_i(p), 1)$ for any point $p \in W_i$ whence it directly follows (since the mapping h_i is a diffeomorphism) that the mapping f_i is an immersion at every point $p \in W_i$.

Now consider a space $\mathbb{R}^N$, where $N = m(n + 1)$. Writing the vectors of the space as $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$, where $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^{n+1}$, we define a mapping

$$f\colon \mathscr{X} \to \mathbb{R}^N$$

putting for any point $p \in \mathscr{X}$

$$f(p) = (f_1(p), \ldots, f_m(p)).$$

The mapping f is clearly smooth.

Let p and q be points of the manifold $\mathscr{X}$ such that $f(p) = f(q)$. Then $f_i(p) = f_i(q)$ for any $i = 1, \ldots, m$ and, in particular, $\varphi_i(p) = \varphi_i(q)$. On the other hand, since charts (W_i, h_i) make up an atlas, there exists an i_0 such that $p \in W_{i_0}$ and hence $\varphi_{i_0}(p) = 1$ (and $f_{i_0}(p) = (h_{i_0}(p), 1)$). Consequently, $\varphi_{i_0}(q) = 1$, which by virtue of property (b) of functions φ_i is possible only for $q \in \overline{W}_i$. Therefore $f_{i_0}(q) = (h_{i_0}(q), 1)$ and hence $h_{i_0}(p) = h_{i_0}(q)$. Since the mapping h_{i_0} is injective, this proves that $p = q$. Thus if $f(p) = f(q)$, then $p = q$, i.e. *the mapping f is injective*. Since the manifold $\mathscr{X}$ is compact, it follows that the mapping f is a moneomorphism.

Now consider the differential df_p of the mapping f at an arbitrary point $p \in \mathscr{X}$. By virtue of the identification $\mathbf{T}_{f(p)}\mathbb{R}^N = \mathbb{R}^N$ we may assume that

$$df_p\colon \mathbf{T}_p\mathscr{X} \to \mathbb{R}^N.$$

But then the differential is obviously made up of mappings f_i in the same way as the mapping f is made up of mappings f_i, i.e. for any vector $A \in \mathbf{T}_p\mathscr{X}$ we have

$$(df)_pA = ((df_1)_pA, \ldots, (df_m)_pA).$$

But if, as above, $p \in W_{i_0}$ (and hence the mapping f_{i_0} is an immersion at a point p), then for any nonzero vector $A \in \mathbf{T}_p\mathscr{X}$ the vector $(df_{i_0})_pA \in \mathbb{R}^{n+1}$ is nonzero. So is therefore the vector $(df)_pA \in \mathbb{R}^N$. This means that the *mapping f is an immersion* (at any point $p \in \mathscr{X}$).

So the mapping f is an immersion which is a moneomorphism. Hence its image $f\mathscr{X}$ is an embedded submanifold of $\mathbb{R}^N$ which is diffeomorphic to the manifold $\mathscr{X}$. ☐

To prove the Whitney theorem (for compact manifolds) it now only remains to prove Proposition 1. To do this we must begin comparatively from afar.

We have already noticed earlier that any compact manifold (which can be covered by a finite family of coordinate neighbourhoods) satisfies the second axiom of countability. More generally, *any manifold covered by a countable system of coordinate neighbourhoods satisfies the second axiom of countability*. Indeed, being homeomorphic to an open set of $\mathbb{R}^n$, every coordinate neighbourhood satisfies the second axiom of countability, and, on the other hand, it is clear that any space with countable covering all elements of which satisfy (in the induced topology) the second axiom of countability obviously satisfies the axiom. ☐

Conversely, it is easy to see that *in any smooth manifold $\mathscr{X}$ which satisfies the second axiom of countability there is a countable base consisting of coordinate neighbourhoods.* Indeed, let $\{O_\alpha\}$ be an arbitrary countable base for $\mathscr{X}$. For any point $p \in \mathscr{X}$ and any coordinate neighbourhood U containing the point p in $\{O_\alpha\}$ there is then an element $O_{\alpha(p,U)}$ such that $p \in O_{\alpha(p,U)} \subset U$. Let O be an arbitrary open submanifold of $\mathscr{X}$. Choose for any point $p \in O$ a coordinate neighbourhood U_p such that $p \in U_p \subset O$ and consider the sets

$$O_p = O_{\alpha\,(p,\,U_p)}.$$

Since $p \in O_p \subset U_p \subset O$ and p ranges over all points of O, we have $\bigcup\limits_p O_p = O$.

Consequently, sets of the form $O_{\alpha\,(p,\,U)}$ (which, by virtue of the inclusion $O_{\alpha(p,\,U)} \subset U$, are the coordinate neighbourhoods) make up a base for the manifold $\mathscr{X}$. To complete the proof it remains to notice that this base is a part of the countable base $\{O_\alpha\}$ and is therefore countable. □

It is known from calculus that many situations become much simpler when we restrict ourselves to points in general position and ignore "sufficiently small" sets. There are at least two approaches to the definition of "sufficiently small" sets of smooth manifolds, a "topological" and a "metric" approach.

The former approach is based on the following general definition.

Definition 2. A subset of a topological space $\mathscr{X}$ is said to be *scattered* (or *nowhere dense*) if its closure has no interior points.

A subset which is the union of a finite or a countable number of scattered sets is called *meager*.

Remark 2. The term "meager set" was introduced by Bourbaki. Meager sets were formerly called "sets of the first category". This is a most unhappy name for many reasons, and it should not be used.

In general, a meager set may be quite "massive". For example, it may coincide with the entire space $\mathscr{X}$ (the field $\mathbb{Q}$ of rational numbers may serve as an example, as in general, any countable topological space having no isolated points).

Definition 3. A topological space $\mathscr{X}$ is said to be a *Baire space* if any meager subset of it has no interior points.

Thus in a Baire space meager sets are indeed meager.

But are smooth manifolds Baire spaces? It turns out that the answer is yes (at least for Hausdorff manifolds), but the proof of the corresponding theorem requires certain preparations for which we have no time. We therefore prefer the other, "metric" approach, and the proof that Hausdorff manifolds are Baire manifolds will be presented in a series of problems.

Definition 4. A topological space $\mathscr{X}$ is said to be *regular* if for any open set $U \subset \mathscr{X}$ and any point $p \in U$ there is an open set V such that $p \in V$ and $\overline{V} \subset U$.

Definition 5. A topological space $\mathscr{X}$ is said to be *locally compact* if each of its points p has a neighbourhood O whose closure $\overline{O}$ is compact.

Problem 1. Prove that a *Hausdorff locally compact space $\mathscr{X}$ is regular.* [**Hint.** By Proposition 3′ of Lecture 9 the compact subpace $\overline{O \cap U}$ is normal. Therefore there is a neighbourhood V of a point p in $\overline{O \cap U}$ such that $\overline{V} \subset O \cap U$.]

Problem 2. Let $\{A_n,\ n \geqslant 1\}$ be a family of nowhere dense subsets of a locally compact Hausdorff (and hence regular) space $\mathscr{X}$, and let U be an arbitrary neighbourhood of some point $p \in \mathscr{X}$. Show that there is a sequence of nonempty open sets U_n, $n \geqslant 0$, such that

(a) the set $\overline{U}_0$ is compact and contained in U,
(b) for any $n \geqslant 1$ the relations

$$\overline{U}_n \subset U_{n-1}, \quad U_n \cap A_n = \varnothing$$

hold. [**Hint.** Since the set A is nowhere dense, there is a nonempty open set V_n in the open set U_{n-1} such that $V_n \cap A_{n-1} = \varnothing$, and since $\mathscr{X}$ is regular, there is a nonempty open set U_n such that $\overline{U}_n \subset V_n$.]

Problem 3. Hence we deduce that *the neighbourhood U contains a point that is not in the union A of sets A_n.* [**Hint.** The sequence $\{\overline{U}_n\}$ is a centred family of closed sets of a compact space $\overline{U}_0$, and the intersection of all sets $\overline{U}_n$ is therefore not empty.]

The statement of Problem 3 means that *any Hausdorff locally compact space is a Baire space.* Since any Hausdorff manifold is obviously locally compact, every Hausdorff manifold is therefore a Baire space.

Recall that a subset A of a Euclidean space $\mathbb{R}^n$ is said to be a *set of measure zero* (or *null set*) if for any $\varepsilon > 0$ there is a finite or countable family of open balls of $\mathbb{R}^n$ which cover A such that the sum of their (n-dimensional!) volumes is less than ε.

In this definition balls may be replaced by parallelepipeds (or even cubes) with edges parallel to the coordinate axes, i.e. by subsets of $\mathbb{R}^n$ whose points $\mathbf{x} = (x^1, \ldots, x^n)$ are characterized by inequalities of the form

$$a^i < x^i < b^i, \quad i = 1, \ldots, n.$$

We shall need the following three properties of null sets known from calculus:

1° *A union* $A = \bigcup_i A_i$ *of a finite or countable family* $\{A_i\}$ *of null sets is a null set.*

2° *For any smooth mapping* $f\colon U \to V$, *where* U *and* V *are open subsets of* $\mathbb{R}^n$, *and any null set* $A \subset U$ *the set* fA *is a null set.*

3° *No null set has interior points.*

To prove Property 1° it suffices to notice that by covering a set A_i for each $i \geqslant 1$ with balls the sum of whose volumes is less than $\varepsilon/2^i$ we obtain a covering of a set A with balls the sum of whose volumes is less than ε.

To prove Property 2°, we first notice that by virtue of Property 1° it suffices to prove only in the additional assumption that there is a closed cube $Q \subset U$ such that A is contained in its interior $\mathring{Q}$. On the other hand, it follows directly from the Lagrange formula applied to functions $f^1, \ldots, f^n$ which specify a mapping f that for any two points $\mathbf{x}, \mathbf{y} \in Q$ we have

$$|f(\mathbf{x}) - f(\mathbf{y})| < M \cdot |\mathbf{x} - \mathbf{y}|,$$

where M is some constant (the maximum of absolute values for first derivatives of functions $f^1, \ldots, f^n$ in the cube Q). Therefore any ball of radius r contained in Q carries the mapping f into a set contained in a ball of radius Mr and hence having a volume at most M^n times larger. By covering the set A with balls of total volume $< \varepsilon/M^n$ we therefore obtain a covering of the set fA with balls of total volume $<\varepsilon$.

To prove Property 3° it suffices to establish that if a finite family $\{Q_1, \ldots, Q_m\}$ of cubes of $\mathbb{R}^n$ with sides parallel to the coordinate axes covers a cube Q, then the sum of the volumes of the cubes $Q_1, \ldots, Q_m$ is less than the volume of Q (and hence bounded below by a positive constant independent of the covering). Let $a_1, \ldots, a_m$ and a be the lengths of the sides of the cubes $Q_1, \ldots, Q_m$ and Q, and $N_1, \ldots, N_m$ and N be the number of points of $\mathbb{R}^n$ with integer coordinates contained in the cubes $Q_1, \ldots, Q_m$ and Q, respectively. Since the cubes $Q_1, \ldots, Q_m$ cover the cube Q, we have

$$N < N_1 + \ldots + N_m.$$

On the other hand, it is clear that

$$[(a-1)^+]^n \leqslant N \leqslant (a+1)^n$$

and similarly

$$[(a_k-1)^+]^n \leqslant N_k \leqslant (a_k+1)^n$$

for any $k = 1, \ldots, m$, where

$$x^+ = \max(x, 0).$$

Therefore

$$[(a-1)^+]^n \leqslant \sum_{k=1}^{m} (a_k+1)^n.$$

Applying this inequality to homothetic cubes $\lambda_{Q_1}, \ldots, \lambda_{Q_m}$ and λ_Q, where $\lambda > 0$, we obtain, with λ sufficiently large,

$$(\lambda a-1)^n \leqslant \sum_{k=1}^{m} (\lambda a_k+1)^n.$$

Hence

$$\left(a-\frac{1}{\lambda}\right)^n \leqslant \sum_{k=1}^{n} \left(a_k+\frac{1}{\lambda}\right)^n,$$

whence, as $\lambda \to +\infty$, we obtain the required inequality for volumes

$$a^n \leqslant \sum_{k=1}^{m} a_k^n.$$

It is astonishing that such an obvious fact should require such elaborate a proof!

We shall also need one, more profound property of null sets.

For any subset $A \subset \mathbb{R}^n$ by a *section associated with* $x^n = t$ we mean a subset A_t of $\mathbb{R}^{n-1}$ consisting of points $\mathbf{x} = (x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}$ such that $(\mathbf{x}, t) = (x_1, \ldots, x_{n-1}, t)$ is in A.

Theorem (Fubini). *Let C be a compact subset of $\mathbb{R}^n$ such that for any $t \in \mathbb{R}$ its section C_t associated with $x_n = t$ is a null set (of $\mathbb{R}^{n-1}$). Then the set C is also a null set.*

Although this theorem is, no doubt, familiar from calculus, we still give its full proof here.

A covering of a closed interval $[0, 1]$ with open intervals is said to be *admissible* if a union of these intervals is contained in the open interval $(-1, 2)$.

Lemma 1. *From any admissible covering of a closed interval* $[0, 1]$ *with open intervals it is possible to single out a finite subcovering the sum of the lengths of whose elements is at most six.*

Proof. We show that any minimal subcovering (i. e. subcovering no element of which can be rejected) has the required property.

Let (a_i, b_i), $1 \leqslant i \leqslant m$, be intervals making up a given minimal subcovering. It is easy to see that $a_i \neq a_j$ for $i \neq j$ (indeed, if $a_i = a_j$, then for $b_i \leqslant b_j$ the interval (a_i, b_i) is unnecessary, and for $b_i \geqslant b_j$, the interval (a_i, b_j)). Indexing, therefore, if necessary, the open intervals, we may assume that $a_i < a_j$ for $i < j$ (and hence $b_i < b_j$ by virtue of minimality). But then $a_i < a_{i+1} < b_i < a_{i+2}$ for any $i = 1, \ldots, m - 2$, because for $a_{i+1} \geqslant b_i$ there would be holes in the covering, and for $b_i \geqslant a_{i+2}$, in view of inequalities $b_i < b_{i+1} < b_{i+2}$ the interval (a_{i+1}, b_{i+1}) would be contained in the union of (a_i, b_i) and (a_{i+2}, b_{i+2}) and hence unnecessary. Consequently, for the sum of the lengths of the intervals of the subcovering under consideration there is an estimate

$$\begin{aligned}(b_1 - a_1) &+ (b_2 - a_2) + (b_3 - a_3) + \ldots + (b_{m-2} - a_{m-2}) \\ &+ (b_{m-1} - a_{m-1}) + (b_m - a_m) \leqslant (a_3 - a_1) \\ &+ (a_4 - a_2) + (a_5 - a_3) + \ldots + (a_m - a_{m-2}) \\ &+ (b_{m-1} - a_{m-1}) + (b_m - a_m) = b_{m-1} + b_m \\ &- a_1 - a_2 < 2\,(b_m - a_1) \leqslant 6\end{aligned}$$

(for in view of admissibility $-1 \leqslant a_1$ and $b_m \leqslant 2$). □

Proof of Fubini's theorem. We may obviously assume without loss of generality that the set C is in the region $0 \leqslant x_n \leqslant 1$. Under the hypothesis, for any $\varepsilon > 0$ and any $t \in [0, 1]$ there are open cubes $Q_t^{(i)}$ in $\mathbb{R}^{n-1}$ that cover the section C_t whose total volume is less than $\varepsilon/6$. Let Q_t be their union (containing C_t), and let Q_t^* be an open subset of $\mathbb{R}^n$ consisting of all points of the form $(\mathbf{x}, t)$, where $\mathbf{x} \in Q_t$ and $0 < t < 1$ (i.e. $Q_t^* = Q_t \times (0, 1)$). Since a closed subset of a compact space is compact, the set $C \setminus Q_t^*$ is compact. On the other hand, the function $f(\mathbf{x}, x^n) = |x_n - t|$, $(\mathbf{x}, x_n) \in \mathbb{R}^n$, is continuous and positive on the set $C \setminus Q_t^*$. Hence there is a number $\alpha > 0$ such that

$$(4) \qquad |x_n - t| > \alpha \text{ for any point } (\mathbf{x}, x_n) \in C \setminus Q_t^*.$$

We may, of course, assume without loss of generality that $\alpha < 1$, i.e. that the covering of a closed interval $[0, 1]$ with intervals $I_t = (t - \alpha, t + \alpha)$ is admissible.

According to Lemma 1 it is possible to choose in this covering a finite subcovering $\{I_{t_1}, \ldots, I_{t_m}\}$ consisting of intervals the sum $2\alpha m$ of whose lengths is less than six.

Let L_{ij} be a parallelepiped of $\mathbb{R}^n$ consisting of points $(\mathbf{x}, \tau)$, $\mathbf{x} \in \mathbb{R}^{n-1}$, $\tau \in \mathbb{R}$ such that $\mathbf{x} \in Q_{t_j}^{(i)}$ and $\tau \in I_{t_j}$ (i.e. $L_{ij} = Q_{t_j}^{(i)} \times I_{t_j}$). It follows directly from formula (4) that parallelepipeds L_{ij} cover the set C. (Indeed, let $(\mathbf{x}, x_n) \in C$. There is a j such that $x_n \in I_{t_j}$, i.e. such that $|x_n - t_j| < \alpha$. By (4), therefore, $(\mathbf{x}, x_n) \in Q^*_{t_j}$, and hence $\mathbf{x} \in Q^{(i)}_{t_j}$ for some i. Consequently, $(\mathbf{x}, x_n) \in L_{ij}$.) On the other hand, the n-dimensional volume of a parallelepiped L_{ij} is obviously equal to the $(n-1)$-dimensional volume of a cube $Q^{(i)}_{t_j}$ multiplied by the length 2α of an interval I_{t_j}. Therefore the total volume of all parallelepipeds L_{ij} is at most $m \cdot \frac{\varepsilon}{6} \cdot 2\alpha < \varepsilon$.

Thus for any $\varepsilon > 0$ the set C allows covering by parallelepipeds of total volume $<\varepsilon$. Hence this set is of measure zero. □

For smooth manifolds null subsets can be naturally defined.

Definition 6. A subset A of a smooth n-dimensional manifold $\mathscr{X}$ is said to be a *null set* if there is a finite or a countable family of charts (U_i, h_i) in $\mathscr{X}$ such that $A \subset \bigcup_i U_i$ and each of the sets $h_i(U_i \cap A)$ is a null set in $\mathbb{R}^n$.

It is clear that *Properties* 1° and 2° *of null sets in* $\mathbb{R}^n$ *remain valid also for null sets in an arbitrary manifold* (in Property 2° one should now understand by U and V any smooth manifolds of the same dimension). In particular, it follows that *for any null set A and any chart (U, h) of $\mathscr{X}$ the set $h(U \cap A)$ is a null set in* $\mathbb{R}^n$, and if the manifold satisfies the second axiom of countability, then, conversely, the *subset $A \subset \mathscr{X}$ is a null set if the set $h(U \cap A)$ is a null set in $\mathbb{R}^n$ for any chart (U, h) of $\mathscr{X}$* (or at least for any chart of an arbitrary countable family of charts whose supports make up a base for the manifold $\mathscr{X}$).

It is useful to compare null sets with meager sets. In general, these two classes of sets are in no way related to each other: there are meager sets that are not null sets (and even have a complete measure, i.e. such that their complement is a null set), and nonmeager null sets (and even null sets with meager complements).

Let, for example, U_i, $1 \leqslant i < \infty$, be a union of a countable system of intervals on the straight line $\mathbb{R}$ whose centres are all possible real points, and the sum of lengths is 2^{-i}. Then the set $A = \bigcap_{i=1}^{\infty} U_i$ is a null set, and since for any i the set $\mathbb{R} \setminus U_i$ is nowhere dense (it is closed and has no interior points), the complement $\mathbb{R} \setminus A = \bigcup_{i=1}^{\infty} (\mathbb{R} \setminus U_i)$ of A is a meager set.

However, since a null set (in $\mathbb{R}^n$ and hence in any $\mathscr{X}$) has no interior points, *any closed null set is nowhere dense.* Every null set that can be represented as a union of a finite or countable number of closed sets (which are necessarily null sets) is therefore a meager set. Such sets have no a generally accepted name. In the absence of a better term we shall call them *null meager sets.*

Lecture 15

The Sard theorem · The analytic part of the proof of the Sard theorem · Direct product of manifolds · The manifold of tangent vectors · Proof of the Whitney theorem

---

Let $f\colon \mathscr{X} \to \mathscr{Y}$ be a smooth mapping of an n-dimensional smooth manifold $\mathscr{X}$ into an m-dimensional smooth manifold $\mathscr{Y}$.

Definition 1. A point $p \in \mathscr{X}$ is said to be a *critical point* of a mapping f if

$$df_p(\mathbf{T}_p\mathscr{X}) \neq \mathbf{T}_q\mathscr{Y}, \quad q = f(p),$$

i.e. if the mapping

$$df_p\colon \mathbf{T}_p\mathscr{X} \to \mathbf{T}_q\mathscr{Y}$$

is not epimorphic (the mapping f is not a submersion at p). A point $q \in \mathscr{Y}$ is said to be the *critical value* of f if there is a critical point $p \in \mathscr{X}$ such that $q = f(p)$.

Comparing this definition with Definition 4 of Lecture 13 we at once see that *critical values are precisely nonregular values of the mapping* f.

Notice that *for* $n < m$ *any point of the manifold* $\mathscr{X}$ *is a critical point of the mapping* f. In this case therefore critical values are points of the set $f\mathscr{X}$, and regular values are points of its complement $\mathscr{Y} \setminus f\mathscr{X}$. (Thus *for* $n < m$ *the inverse image of any regular value is empty.*)

Now we can formulate the main theorem of this lecture:

Theorem 1. *If a manifold* $\mathscr{X}$ *satisfies the second axiom of countability, then for any smooth mapping* $f\colon \mathscr{X} \to \mathscr{Y}$

the set $C(f)$ of its critical values is a null set, and if a manifold $\mathscr{Y}$ is Hausdorff, even a null meager set.

But if the manifold $\mathscr{X}$ is compact (and the manifold $\mathscr{Y}$ is Hausdorff), then the set $C(f)$ is closed and nowhere dense.

This theorem is usually called the **Sard theorem** although it was proved by Brown prior to Sard and by Dubovitsky independently of Sard.

Corollary. *If* $\dim \mathscr{X} < \dim \mathscr{Y}$, *then every smooth mapping* $f\colon \mathscr{X} \to \mathscr{Y}$ *is trivially not surjective (the set* $\mathscr{Y} \setminus f\mathscr{X}$ *is not empty).* □

We shall need only this corollary to prove the Whitney theorem.

We emphasize that smooth manifolds of class C^∞ or C^ω are meant in Theorem 1. However, looking through its proof given below we can see that it is also valid for C^r-manifolds, where $r \geqslant n - m + 2$ for $n \geqslant m$ and $r \geqslant 2$ for $n \leqslant m$. [Moreover, resorting to some technical tricks this r can be reduced by yet another unity. Appropriate examples show that it is impossible to reduce r still more.]

We derive the Sard theorem from the following proposition.

Proposition 1. *Let* $f\colon U \to \mathbb{R}^m$ *be a smooth mapping of an open set* $U \subset \mathbb{R}^n$ *into* $\mathbb{R}^m$, *and let* K *be an arbitrary compact subset of the set of critical points of* f. *Then the set* fK *is a null set.*

Notice that since $\mathbb{R}^n$ is a Hausdorff space, the set fK is closed and hence nowhere dense. On the other hand, the set of critical points of f is obviously closed and hence a union of a countable family of compact subsets. Consequently, the set $C(f)$ of critical points of f is a null meager set.

We thus see that in the special case $\mathscr{X} = U$, $\mathscr{Y} = \mathbb{R}^m$ the Sard theorem is a direct consequence of Proposition 1. It turns out that in the general case it can be easily reduced to this proposition.

Indeed, for any open subset O of a manifold $\mathscr{X}$ the critical points of f contained in O are obviously nothing but the critical points of its restriction $f|_O$ on O. Hence

$$C(f) = \bigcup_\alpha C(f|_{U_\alpha})$$

for any countable base $\{U_\alpha\}$ for $\mathscr{X}$.

If every U_α is the support of some chart (U_α, f_α) (which, as we know, can always be attained) and if $f(U_\alpha) \subset V_\alpha$, where (V_α, k_α) is a chart of a manifold $\mathcal{Y}$, then $C(f \mid_{U_\alpha})$ is the image under the diffeomorphism k_α^{-1} of a set $C(g_\alpha)$, where $g_\alpha = k_\alpha \circ (f \mid_{U_\alpha}) \circ h_\alpha^{-1}: U_\alpha \to \mathbb{R}^m$.

For the same reasons as earlier, every set

$$C(f \mid_{U_\alpha}) = k_\alpha^{-1} C(g_\alpha)$$

is a union of a countable family of compact null sets of the form fK, where K is some compact set consisting of critical points of the mapping f.

Consequently, a union of a countable family of such sets is the entire set $C(f)$. Since a union of a countable family of null sets is a null set, this proves that the set $C(f)$ is a null set.

If a manifold $\mathcal{Y}$ is Hausdorff, then all sets fK are closed, and hence the set $C(f)$ is a null meager set.

Finally, if in addition the manifold $\mathcal{X}$ is compact, then the set of all critical points of the mapping f can be decomposed as a union $K_1 \cup \ldots \cup K_N$ of a finite number of compact sets each of which is contained in one of the coordinate neighbourhoods U_α. In this case therefore $C(f) = fK_1 \cup \ldots \cup fK_N$, where, $\mathcal{Y}$ being a Hausdorff manifold, all sets fK_i are closed and hence nowhere dense. Consequently, the set $C(f)$ is also closed and nowhere dense. □

It thus remains to prove Proposition 1.

First notice that *Proposition* 1 *is trivially proved for* $n < m$. Indeed, assuming in this case that $\mathbb{R}^m = \mathbb{R}^n \times \mathbb{R}^{m-n}$ we can introduce into consideration an open set $U \times \mathbb{R}^{m-n}$ of $\mathbb{R}^m$ and its smooth mapping

$$F: U \times \mathbb{R}^{m-n} \to \mathbb{R}^m,$$

which is a composition of a projection onto the first factor $U \times \mathbb{R}^{m-n} \to U$ and of the mapping $f: U \to \mathbb{R}^m$. For a compact set $K \subset U$ we have

$$f(K) = F(K \times 0),$$

where $K \times 0$ is a subset of $\mathbb{R}^m$ consisting of points of the form $(x, 0)$, where $\mathbf{x} \in K$, and $\mathbf{0}$ is a zero vector of $\mathbb{R}^{m-n}$.

On the other hand, it is clear that the set $K \times 0$ is a null set of $\mathbb{R}^m$ (since it can be covered by a finite system of parallelepipeds of arbitrarily small height). According to Property 2° of null sets of Lecture 14 therefore the set $F(K \times 0)$ is also a null set. □

This proves the Sard theorem for $n < m$ and hence the above corollary (which we need to prove the Whitney theorem).

Now we proceed to a general discussion of Proposition 1. Although this proposition belongs, strictly speaking, to calculus, we shall still prove it here despite the fact that the proof is somewhat awkward and tiresome. (Those interested in the Whitney theorem may skip the proof for the time being.)

Let a mapping $f\colon U \to \mathbb{R}^m$ be given by functions $f^1, \ldots, f^m\colon U \to \mathbb{R}$, and let C_k, where $k \geqslant 1$, be the set of all points of U at which all partial derivatives of $f^1, \ldots, f^m$ of orders $\leqslant k$ are zero. It is clear that

$$C_0 \supset C_1 \supset \ldots \supset C_k \supset \ldots,$$

where C_0 is the set of all critical points of $f\Big($i.e. points in which the rank of the matrix $\left\| \dfrac{\partial f^j}{\partial x^i} \right\|$ is less than $m\Big)$.

Lemma 1. *If* $C_{s+1} = \varnothing$ *or* $s > \dfrac{n}{m} - 1$, *then for any compact set* $K \subset C_s$ *the set* fK *is of measure zero.*

Proposition 1 is derived from the lemma by simple induction.

Indeed, there is an s such that the set $f(K \cap C_s)$ is of measure zero $\left(\text{according to the lemma, any } s > \dfrac{n}{m} - 1 \text{ suits}\right)$. On the other hand, $K \cap C_0 = K$. To prove Proposition 1 it is therefore sufficient to establish that *if for some* $s \geqslant 1$ *the set* $f(K \cap C_s)$ *is of measure zero, then the set* $f(K \cap C_{s-1})$ *is also of measure zero.*

Let K_r, $r \geqslant 1$, be a subset of the set $K \cap C_{s-1}$ that consists of points $\mathbf{x} \in K \cap C_{s-1}$ such that $|\mathbf{x} - \mathbf{y}| \geqslant 1/r$ for any point $\mathbf{y} \in K \cap C_s$. Being a closed subset of a compact set K the set K_s is compact.

It is clear that

$$K \cap C_{s-1} = (K \cap C_s) \cup \bigcup_{r=1}^{\infty} K_r.$$

We therefore need only to prove that *for any* $r \geqslant 1$ *the set* fK_r *is of measure zero.*

To this end consider an open set $U^* = U \setminus C_s$. Let $f^* = f|_{U^*}$, and let C_k^* be sets C_k constructed for the function f^*. Clearly, $C_k^* = C_k \setminus C_s$, and hence $C_s^* = \varnothing$ and $K_r \subset C_{s-1}^*$. Lemma 1 is therefore applicable to the mapping f^* and the compact set K_r. Consequently, the set fK_r is indeed of measure zero. □

Thus it only remains to prove Lemma 1.

Proof of Lemma 1. The lemma consists of two statements with different premises, but with the same inference. We first consider the first statement (with the premise $C_{s+1} = \varnothing$). Since a set K is compact, to prove the statement it suffices to find for any point $\mathbf{x}_0 \in K$ a neighbourhood $V \subset U$ such that $f(K \cap V)$ is of measure zero. To this end we find a neighbourhood V of $\mathbf{x}_0$ such that its closure $\overline{V}$ is compact and the set $f(C_s \cap \overline{V})$ is of measure zero. Then the set $f(K \cap V) \subset f(C_s \cap \overline{V})$ is also of measure zero.

We proceed by induction on n. Since for $n = 0$ the statement is obviously true, we need only to prove that if it is true for $n - 1$, it is also true for n.

In view of the condition $C_{s+1} = \varnothing$ there is a nonzero derivative of order $s + 1$ of one of the functions $f^1, \ldots, f^m$ at $\mathbf{x}_0$. Interchanging, if necessary, the coordinates, we may assume without loss of generality that the derivative is of the form $\frac{\partial \varphi}{\partial x_n}$, where φ is some derivative of order s.

Let g be a mapping $U \to \mathbb{R}^m$ defined as

$$g(\mathbf{x}, x_n) = (\mathbf{x}, \varphi(\mathbf{x}, x_n)), \quad (\mathbf{x}, x_n) \in U.$$

The Jacobian of this mapping is equal to $\frac{\partial \varphi}{\partial x_n}$ and hence is nonzero at $\mathbf{x}_0$. Consequently, there is a neighbourhood O' of $\mathbf{x}_0$ such that g is a diffeomorphism of the neighbourhood onto some open set O, and hence a mapping

$$h\colon O \xrightarrow{(g|_{O'})^{-1}} O' \xrightarrow{f|_{O'}} \mathbb{R}^m$$

is defined.

Let V be a neighbourhood of a point $\mathbf{x}_0$ such that $\overline{V}$ is contained in O' and compact. We show that the *set* $f(C_s \cap \overline{V})$ *is of measure zero.*

Two cases should be considered here: $s = 0$ and $s > 0$. We first consider the case $s > 0$ as a simpler one.

For $s > 0$ the function φ, being the derivative of order s of one of the functions $f^1, \ldots, f^m$, is zero on C_s. Hence the set $g(C_s)$ is in the subspace $\mathbb{R}^{n-1} \times \{0\}$ of $\mathbb{R}^n$ defined by the equation $x_n = 0$, and therefore

$$f(C_s \cap \overline{V}) = h_0(g(C_s \cap \overline{V})),$$

where h_0 is the restriction of h to $(\mathbb{R}^{n-1} \times \{0\}) \cap O$. This is an obvious proof, since the mapping h_0 is the mapping of an open set of $\mathbb{R}^{n-1}$, and hence under the induction hypothesis (applied to the mapping h_0 and the compact set $g(C_s \cap \overline{V})$) the set $h_0(g(C_s \cap \overline{V}))$ is of measure zero.

For $s = 0$ the role of the derivative φ is played by one of the functions $f^1, \dots, f^m$ (nonzero on the set C_0). Interchanging, if necessary, the coordinates in R^m, we may assume without loss of generality that $\varphi = f^m$.

Let O_t be a section associated with $x_n = t$ of the set O (the set of all points $\mathbf{x} \in \mathsf{R}^{n-1}$ for which the point $(\mathbf{x}, t) \in \mathsf{R}^n$ belongs to O).

If $\mathbf{x} \in O_t$ and $g(\mathbf{x}, a) = (\mathbf{x}, t)$, then by definition $f^m(\mathbf{x}, a) = \varphi(\mathbf{x}, a) = t$, and therefore

$$h(\mathbf{x}, t) = f(\mathbf{x}, a) = (f^1(\mathbf{x}, a), \dots, f^{m-1}(\mathbf{x}, a), t).$$

This means that putting

$$h_t(\mathbf{x}) = (f^1(\mathbf{x}, a), \dots, f^{m-1}(\mathbf{x}, a)),$$

where $\mathbf{x} \in O_t$ and $g(\mathbf{x}, a) = (\mathbf{x}, t)$ (i.e. $t = f^m(\mathbf{x}, a)$), we obtain a smooth mapping

$$h_t \colon O_t \to \mathsf{R}^{m-1}$$

such that

$$(1) \qquad h(\mathbf{x}, t) = (h_t(\mathbf{x}), t)$$

for any point $(\mathbf{x}, t) \in O$.

It follows from formula (1) that the Jacobian matrix J of the mapping h is of the form

$$\left\|\begin{array}{ccc|c} & & & * \\ & J_t & & \vdots \\ & & & * \\ \hline 0 & \dots & 0 & 1 \end{array}\right\|,$$

where J_t is the Jacobian matrix of h_t.

The rank of J at $(\mathbf{x}, t)$ is therefore less than m if and only if that of J_t at $\mathbf{x}$ is less than $m - 1$. But by definition the rank of J_t at $\mathbf{x} \in O_t$ is less than $m - 1$ if and only if $\mathbf{x}$ is a critical point of h_t, and similarly the rank of J at $(\mathbf{x}, t) \in O$ is less than m if and only if this point is a critical point of the mapping h (and hence the point $(\mathbf{x}, a) = g^{-1}(\mathbf{x}, t) \in O'$ is a critical point of the mapping f, i.e. is in the set $C_0 \cap O'$). Consequently, the section $g(C_0 \cap O')_t$ of the set $g(C_0 \cap O')$ is nothing but the set of critical points of h_t. Therefore under the induction hypothesis (applied to the mapping h_t and the compact set $g(C_0 \cap O')_t \cap g(\overline{V})_t)$) the subset

$$h_t(g(C_0 \cap O')_t \cap g(\overline{V})_t)$$

of R^{n-1} is a null set.

Since this last set is by (1) a section associated with $x_m = t$ of the set

$$h(g(C_0 \cap O') \cap g(\overline{V})) = (h \circ g)(C_0 \cap \overline{V}) = f(C_0 \cap \overline{V}),$$

we thus obtain that all sections $f(C_0 \cap \overline{V})_t$ of the set $f(C_0 \cap \overline{V})$ are null sets. But then by the Fubini theorem the set $f(C_0 \cap \overline{V})$ is also a null set. This completes the proof of Lemma 1 for $C_{s+1} = \varnothing$.

Now let $s > \frac{n}{m} - 1$. Since being compact the set K can be covered with a finite number of cubes of $\mathbb{R}^n$, it suffices to prove that for any cube $Q \subset \mathbb{R}^n$ the set $f(C_s \cap Q)$ is of measure zero.

Let a be an edge of Q, and let $k \geqslant 1$. Dividing Q by planes parallel to its faces into k^n cubes with edges a/k, consider one of the cubes Q' which intersects with C_s.

Since all derivatives of order $s+1$ of the functions $f^1, \ldots, f^m$ are bounded in Q, it follows immediately from the Taylor formula applied to these functions that for any points $\mathbf{x} \in C_s \cap Q$ and $\mathbf{y} \in Q$ we have

$$| f(\mathbf{y}) - f(\mathbf{x}) | \leqslant M \cdot | \mathbf{y} - \mathbf{x} |^{s+1},$$

where M is some constant number. Since the diameter of Q' is obviously equal to $\frac{a}{k}\sqrt{n}$, it follows that the diameter of its image fQ' under the mapping f is at most $M\left(\frac{a}{k}\sqrt{n}\right)^{s+1}$, and hence this image is contained in a cube of $\mathbb{R}^m$ with edge $2M\left(\frac{a}{k}\sqrt{n}\right)^{s+1}$ and volume

$$\left[2M\left(\frac{a}{k}\sqrt{n}\right)^{s+1}\right]^m = \frac{B}{k^{(s+1)m}},$$

where $B = (2Ma^{s+1}(\sqrt{n})^{s+1})^m$ is a constant independent of k.

Since there are at most k^n cubes Q' and their union contains the set $C_s \cap Q$, it follows that the set $f(C_s \cap Q)$ is contained in the union of the cubes whose total volume does not exceed the number

$$\frac{B}{k^{(s+1)m}} k^n = \frac{B}{k^{(s+1)m-n}}$$

and hence tends to zero as $k \to \infty$ (recall that under the hypothesis $(s+1)m - n > 0$).

The set $f(C_s \cap Q)$ can thus be covered with (even a finite) family of cubes the sum of whose volumes is arbitrarily small. Hence this set is of measure zero.

This completes the proof of Lemma 1 and hence of the Sard theorem. □

The Sard theorem (or more exactly the corollary to it) is the key to the proof of Proposition 1 of the preceding lecture (and hence of the Whitney theorem), but to apply this key we need some simple but interesting in their own right constructions.

Let $\mathcal{X}$ and $\mathcal{Y}$ be two manifolds (of dimension n and m, respectively), and let $\mathcal{X} \times \mathcal{Y}$ be the set of all pairs (p, q), where $p \in \mathcal{X}$, $q \in \mathcal{Y}$. For any sets $U \subset \mathcal{X}$ and $V \subset \mathcal{Y}$ the set $U \times V$ is a subset of $\mathcal{X} \times \mathcal{Y}$, and for any mappings h: $U \to \mathbb{R}^n$ and k: $V \to \mathbb{R}^m$ the formula

$$(h \times k)(p, q) = (h(p), k(q))$$

defines some mapping

$$h \times k\colon\ U \times V \to \mathbb{R}^{n+m}$$

(we identify $\mathbb{R}^n \times \mathbb{R}^m$ with $\mathbb{R}^{n+m}$ here). If mappings h and k are injective, then the mapping $h \times k$ is also injective, and if the sets $h(U)$ and $k(V)$ are open (in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively), then the set

$$(h \times k)(U \times V) = h(U) \times k(V)$$

is open in $\mathbb{R}^{n+m}$. This means that if (U, h) and (V, k) are the charts, then $(U \times V, h \times k)$ is also a chart. Moreover it is easy to see that *if the chart (U, h) is compatible with the chart (U', h') and (V, k) with (V', k'), then the chart $(U \times V, h \times k)$ is compatible with the chart $(U' \times V', h' \times k')$.* Indeed, it is clear that

$$(U \times V) \cap (U' \times V') = (U \cap U') \times (V \cap V')$$

and similarly

$$\begin{aligned}(h(U) \times k(V)) &\cap (h'(U') \times k'(V'))\\ &= (h(U) \cap h'(U')) \times (k(V) \cap k'(V'))\end{aligned}$$

(we agree to assume that $A \times B = \varnothing$ if $A = \varnothing$ or $B = \varnothing$), with

$$\begin{aligned}&(h \times k\,|_{(U\times V)\cap(U'\times V')}) \circ (h' \times k'\,|_{(U\times V)\cap(U'\times V')})^{-1}\\ &= [(h|_{U\cap U'}) \circ (h'\,|_{U\cap U'})^{-1}] \times [(k\,|_{V\cap V'}) \circ (k'\,|_{V\cap V'})^{-1}],\end{aligned}$$

and to complete the proof we should notice that for any diffeomorphisms φ: $W \to W_1$ and φ': $W' \to W_1'$ of open sets of $\mathbb{R}^n$ and $\mathbb{R}^m$ the mapping

$$\varphi \times \varphi'\colon W \times W' \to W_1 \times W_1'$$

is also a diffeomorphism of open sets (of the space $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$). □

Thus charts $(U \times V, h \times k)$ constructed for all possible charts (U, k) and (V, l) of manifolds $\mathscr{X}$ and $\mathscr{Y}$ make up an atlas on $\mathscr{X} \times \mathscr{Y}$.

Definition 2. The corresponding smoothness on $\mathscr{X} \times \mathscr{Y}$ is called a *direct product* of the smoothnesses of $\mathscr{X}$ and $\mathscr{Y}$, and the set $\mathscr{X} \times \mathscr{Y}$ provided with this smoothness is called a *direct product* of manifolds $\mathscr{X}$ and $\mathscr{Y}$. Its dimension is equal to the sum of dimensions of the cofactors:

$$\dim (\mathscr{X} \times \mathscr{Y}) = \dim \mathscr{X} + \dim \mathscr{Y}.$$

The topology of the manifold $\mathscr{X} \times \mathscr{Y}$ is obviously a direct product of the topologies of $\mathscr{X}$ and $\mathscr{Y}$ (see Lecture 8).

Remark 1. For an arbitrary group $\mathscr{G}$ we can define a mapping

$$\mathscr{G} \times \mathscr{G} \to \mathscr{G}, \ (a, b) \mapsto ab, \ a, b \in \mathscr{G}. \tag{2}$$

A group $\mathscr{G}$, which is a smooth manifold, for which mapping (2) is smooth is said to be a *Lie group* (the term *smooth group* is also used). For any matrix Lie group $\mathscr{G}$ (see Definition 1 of Lecture 11) the smoothness constructed in Proposition 1 of Lecture 11 has the property (check it!) that with respect to this smoothness the group $\mathscr{G}$ is a Lie group. This justifies our terminology.

Remark 2. In Semester IV we prove that if a subgroup $\mathscr{G}$ of the group GL $(n, \mathbb{R})$ is an embedded submanifold (and hence a Lie group), then it is a matrix Lie group in the sense of Definition 1 of Lecture 11. This shows that Definition 1 of Lecture 11 is not, as it may seem, an *ad hoc* definition and introduces an entirely natural notion.

When $\mathscr{X} = \mathscr{Y}$ there arises a manifold $\mathscr{X} \times \mathscr{X}$ which is called a *square* of a manifold $\mathscr{X}$. A subset Δ of the manifold that consists of points of the form (p, p), $p \in \mathscr{X}$ is called the *diagonal* of $\mathscr{X}$.

It is easy to see that a *manifold $\mathscr{X}$ (or more generally a topological space $\mathscr{X}$) is Hausdorff if and only if the diagonal Δ is closed in $\mathscr{X} \times \mathscr{X}$*. Indeed, the condition that distinct points p and q should have nonintersecting neighbourhoods U and V means exactly that the neighbourhood $U \times V$ of a point $(p, q) \in \mathscr{X} \times \mathscr{X}$ does not intersect with Δ. □

For any Hausdorff manifold $\mathscr{X}$ therefore a manifold $\mathscr{X} \times \mathscr{X} \setminus \Delta$ is defined, the dimension of which is $2n$, where n is the dimension of $\mathscr{X}$.

Another construction we need associates an arbitrary n-dimensional manifold $\mathscr{X}$ with some new $2n$-dimensional manifold $\mathbf{T}\mathscr{X}$.

Let

$$\mathbf{T}\mathscr{X} = \bigsqcup_{p \in \mathscr{X}} \mathbf{T}_p\mathscr{X}$$

be a disjoint union of all subspaces $\mathbf{T}_p\mathscr{X}, p \in \mathscr{X}$. [Points of the set $\mathbf{T}\mathscr{X}$ are thus all possible tangent vectors A of $\mathscr{X}$.] For every vector $A \in \mathbf{T}\mathscr{X}$ (the only!) point $p \in \mathscr{X}$ for which $A \in \mathbf{T}_p\mathscr{X}$ is designated πA. This gives rise to a mapping

$$\pi\colon \mathbf{T}\mathscr{X} \to \mathscr{X}$$

with the property that $\pi^{-1}(p) = \mathbf{T}_p\mathscr{X}$ for every point $p \in \mathscr{X}$.

For an arbitrary open subset $U \subset \mathscr{X}$ the subset $\pi^{-1}U \subset \mathbf{T}\mathscr{X}$ can be naturally identified with a set $\mathbf{T}U$. If U is the support of a chart $(U, h) = (U, x^1, \ldots, x^n)$, a mapping

$$\mathbf{T}h\colon \mathbf{T}U \to \mathbb{R}^{2n}$$

is defined which carries an arbitrary vector $A \in \mathbf{T}U$ into a vector

$$(\mathbf{T}h)\, A = (x^1, \ldots, x^n, a^1, \ldots, a^n) \in \mathbb{R}^{2n},$$

where $x^1, \ldots, x^n$ are the coordinates of a point $p = \pi A$ in a chart (U, h) and $a^1, \ldots, a^n$ are the coordinates of a vector A in a basis

$$\left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^n}\right)_p$$

for a space $\mathbf{T}_p\mathscr{X}$. (Thus

$$A = a^1 \left(\frac{\partial}{\partial x^1}\right)_p + \ldots + a^n \left(\frac{\partial}{\partial x^n}\right)_p.$$

Notice that the numbers $a^1, \ldots, a^n$ are defined uniquely by a vector A). Clearly, $\mathbf{T}h$ is a bijective mapping, and hence the pair $(\mathbf{T}U, \mathbf{T}h)$ is a chart in $\mathbf{T}\mathscr{X}$.

Let (U, h) and (U', h') be two charts (intersecting, for definiteness) of a manifold $\mathscr{X}$ and let

$$x^{i'} = x^{i'}(x^1, \ldots, x^n), \quad i = 1, \ldots, n, \tag{3}$$

be transition formulas for the corresponding local coordinates (in the intersection $U \cap U'$). By definition (see formula (1) of Lecture 12) for every point $p \in U \cap U'$ the coordinates $a^1, \ldots, a^n$ and $a^{1'}, \ldots, a^{n'}$ of an arbitrary vector $A \in \mathbf{T}_p\mathscr{X}$ in charts (U, h) and (U', h') are related by

$$a^{i'} = \left(\frac{\partial x^{i'}}{\partial x^i}\right)_p a^i, \tag{4}$$

where $\left(\frac{\partial x^{i'}}{\partial x^i}\right)_p$ are values of partial derivatives of functions (3) at a point p.

On the other hand, it is clear that $\mathbf{T}h$ and $\mathbf{T}h'$ map a set

$$\mathbf{T}(U \cap U') = \mathbf{T}U \cap \mathbf{T}U'$$

onto open sets $h(U \cap U') \times \mathbb{R}^n$ and $h'(U \cap U') \times \mathbb{R}^n$ of $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$, respectively, formulas (3) and (4) giving a mapping $\mathbf{T}h' \circ (\mathbf{T}h)^{-1}$ of the former set onto the latter. Since the Jacobian det $\left\| \left(\frac{\partial x^{i'}}{\partial x^i}\right)_p \right\|$ is nonzero at all points $p \in U \cap U'$, it follows that the mapping is a diffeomorphism.

The charts $(\mathbf{T}U, \mathbf{T}h)$ *and* $(\mathbf{T}U', \mathbf{T}h')$ *are thus compatible*. This conclusion clearly remains valid for $U \cap U' = \varnothing$ as well (for if $U \cap U' = \varnothing$, then $\mathbf{T}U \cap \mathbf{T}U' = \varnothing$).

Thus charts of the form $(\mathbf{T}U, \mathbf{T}h)$ make up an atlas on $\mathbf{T}\mathscr{X}$ and hence define some smoothness on $\mathbf{T}\mathscr{X}$.

Definition 3. The constructed smooth manifold $\mathbf{T}\mathscr{X}$ is called a *manifold of tangent vectors* of a manifold $\mathscr{X}$. Its dimension is equal to $2n$, where $n = \dim \mathscr{X}$.

Notice that because of the derivatives in (4) the smoothness class of $\mathbf{T}\mathscr{X}$ is a unity less than the smoothness class

r of $\mathscr{X}$ (for $r = \infty$ or $r = \omega$ the smoothness class clearly remains unchanged).

By definition the numbers $x^1, \ldots, x^n, a^1, \ldots, a^n$ are the local coordinates corresponding to a chart ($\mathrm{T}U$, $\mathrm{T}h$). (Thus symbols $x^1, \ldots, x^n$ designate both local coordinates in U and a part of the local coordinates in $\mathrm{T}U$. No confusion arises on careful consideration.)

In local coordinates $x^1, \ldots, x^n, a^1, \ldots, a^n$ (on $\mathrm{T}U$) and $x^1, \ldots, x^n$ (on U) a mapping π can be written as follows:

$$x^i = x^i, \; i = 1, \ldots, n,$$

where on the left x^i are the coordinates on U and on the right on $\mathrm{T}U$. We thus see that the *mapping π is smooth and is a submersion*. By the general Proposition 1 of Lecture 13 *every tangent space $\mathrm{T}_p\mathscr{X} = \pi^{-1}(p)$ is therefore an embedded submanifold of a manifold $\mathrm{T}\mathscr{X}$*. Numbers $a^1, \ldots, a^n$ are the coordinates on that submanifold (defined on the entire $\mathrm{T}_p\mathscr{X}$).

The manifold $\mathrm{T}\mathscr{X}$ has a subset $\mathscr{X}_0$ consisting of zero vectors of spaces $\mathrm{T}_p\mathscr{X}$.

It is easy to see that *$\mathscr{X}_0$ is a closed submanifold diffeomorphic to a manifold $\mathscr{X}$* (the diffeomorphism $\mathscr{X}_0 \to \mathscr{X}$ is induced by a mapping π: $\mathrm{T}\mathscr{X} \to \mathscr{X}$).

The set $\mathrm{T}\mathscr{X} \setminus \mathscr{X}_0$ is therefore open and hence is a manifold. The dimension of that manifold is $2n$, where $n = \dim \mathscr{X}$.

Now we are in a position to prove Proposition 1 of Lecture 14 (and hence the Whitney theorem).

Proof of Proposition 1 of Lecture 14. Let a manifold $\mathscr{X}$ be embedded in $\mathbb{R}^N$. For any two distinct points p, $q \in \mathscr{X}$ (i.e. an arbitrary point $(p, q) \in \mathscr{X} \times \mathscr{X} \setminus \Delta$) a straight line through these points is defined in $\mathbb{R}^N$. Let $f_1(p, q)$ be a one-dimensional subspace of $\mathbb{R}^N$ associated with that straight line (i.e. a parallel straight line through a point 0). Since one-dimensional subspaces of $\mathbb{R}^N$ make up an $(N-1)$-dimensional projective space $\mathbb{R}P^{N-1}$, this yields a mapping

$$f_1 \colon \mathscr{X} \times \mathscr{X} \setminus \Delta \to P^{N-1}. \tag{5}$$

Similarly we can define a mapping

$$f_2\colon \mathbf{T}\mathscr{X}\setminus\mathscr{X}_0 \to \mathbb{R}P^{N-1} \tag{6}$$

by taking as $f_2(A)$ for any vector $A \in \mathbf{T}\mathscr{X}\setminus\mathscr{X}_0$ a one-dimensional subspace of $\mathbb{R}^N$ (a point of $\mathbb{R}P^{N-1}$) generated by that vector (or more precisely, by a vector $(d\iota_p)A \in \mathbf{T}_p\mathbb{R}^N = \mathbb{R}^N$, where $d\iota_p$ is the differential of an embedding $\iota\colon \mathscr{X} \to \mathbb{R}^N$ at a point $p = \pi(A)$).

Problem 1. Prove that mappings (5) and (6) are smooth. [**Hint.** Write these mappings in local coordinates.]

Now let

$$\mathscr{Y} = (\mathscr{X}\times\mathscr{X}\setminus\Delta) \sqcup (\mathbf{T}\mathscr{X}\setminus\mathscr{X}_0)$$

be a disjoint union of manifolds $\mathscr{X}\times\mathscr{X}\setminus\Delta$ and $\mathbf{T}\mathscr{X}\setminus\mathscr{X}_0$. This union is naturally a $2n$-dimensional smooth manifold, and its mapping

$$f\colon \mathscr{Y} \to \mathbb{R}P^{N-1}$$

coinciding on $\mathscr{X}\times\mathscr{X}\setminus\Delta$ with f_1 and on $\mathbf{T}\mathscr{X}\setminus\mathscr{X}_0$ with f_2 is a smooth mapping. If therefore $2n < N - 1$, then by the corollary to the Sard theorem the mapping f is trivially not surjective. This proves that *in* $\mathbb{R}^N$ *there is a one-dimensional subspace* L *which is neither in the image of* f_1 *nor in that of* f_2.

Now consider a projection

$$\mathbb{R}^N \to L^\perp \tag{7}$$

of $\mathbb{R}^N$ parallel to a straight line L onto its orthogonal complement $L^\perp$(which is an $(N-1)$-dimensional subspace of $\mathbb{R}^N$). Choosing a basis in $L^\perp$ we may assume the projection $\mathbb{R}^N \to \mathbb{R}^{N-1}$ to be a mapping. Let

$$\rho\colon \mathscr{X} \to \mathbb{R}^{N-1} \tag{8}$$

be a restriction of that mapping to $\mathscr{X} \subset \mathbb{R}^N$. Since projection (7) is continuous and open, its restriction (8) is also continuous and open. On the other hand, the statement that $L \notin \operatorname{Im} f_1$ precisely means that every straight line parallel to a straight line L intersects a submanifold $\mathscr{X}$ at most at one point, i.e. that mapping (8) is injective. Being injective continuous and open, mapping (8) is, consequently, a moneomorphism.

Now consider the condition that $L \notin \operatorname{Im} f_2$ (which geometrically means that a straight line through a point p parallel to a straight line L is not tangent at that point to a submanifold $\mathscr{X}$). The differential $(d\rho)_p$ of mapping (8) at every point $p \in \mathscr{X}$ is obviously a restriction to $\mathbf{T}_p\mathscr{X} \subset \mathbb{R}^N$ of the differential of projection (7). Since in view of mapping (7) being linear its differential coincides with the mapping itself, it follows that the differential $(d\rho)_p$ makes only vectors of $\mathbf{T}_p\mathscr{X} \cap L$ vanish. But this intersection either consists of only a zero vector (when $L \not\subset \mathbf{T}_p\mathscr{X}$) or coincides with L (when $L \subset \mathbf{T}_p\mathscr{X}$). If therefore $L \not\subset \mathbf{T}_p\mathscr{X}$, then mapping (8) is an immersion at p. Since the inclusion $L \subset \mathbf{T}_p\mathscr{X}$ precisely means that $L = f_2(p)$, this proves that the condition $L \notin \operatorname{Im} f_2$ is equivalent to the fact that the projection is an immersion (at every point $p \in \mathscr{X}$) and hence a diffeomorphism onto its own image.

The image $\rho(\mathscr{X})$ of a manifold under a projection ρ is therefore an embedded submanifold of $\mathbb{R}^{N-1}$ diffeomorphic to a manifold $\mathscr{X}$. The manifold $\mathscr{X}$ can be embedded in $\mathbb{R}^{N-1}$. $\square$

Problem 2. Show that any n-dimensional manifold that can be embedded in $\mathbb{R}^n$ can be immersed in $\mathbb{R}^{2n}$.

Lecture 16

Tensors · Tensor fields · Vector fields and differentiations · Lie algebra of vector fields

---

Recall (see Lecture II.5) that by *tensor S, of type (a, b)* where $a \geqslant 0$; $b \geqslant 0$, on a vector space $\mathscr{V}$ is meant a mapping that associates an arbitrary basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ for $\mathscr{V}$ with a collection of n^{a+b} numbers $S^{j_1 \ldots j_b}_{i_1 \ldots i_a}$ called the *components* of the tensor S in that basis and having the property that for any two bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{e}_{1'}, \ldots, \mathbf{e}'_n$ for $\mathscr{V}$ the corresponding components $S^{j_1 \ldots j_b}_{i_1 \ldots i_a}$ and $S^{j'_1 \ldots j'_b}_{i'_1 \ldots i'_a}$ are related by the formula

$$S^{j'_1 \ldots j'_b}_{i'_1 \ldots i'_a} = c^{i_1}_{i'_1} \ldots c^{i_a}_{i'_a} c^{j'_1}_{j_1} \ldots c^{j'_b}_{j_b} S^{j_1 \ldots j_b}_{i_1 \ldots i_a},$$

where $c^{i'}_i$ and $c^i_{i'}$ are the components of the inverse transition matrices, i.e. numbers such that $\mathbf{e}_{i'} = c^i_{i'}\mathbf{e}_i$ and $\mathbf{e}_i = c^{i'}_i \mathbf{e}_{i'}$. Every tensor correctly defines a multilinear functional

$$S(\mathbf{x}_1, \ldots, \mathbf{x}_a, \xi^1, \ldots, \xi^b)$$
$$= S^{j_1 \ldots j_b}_{i_1 \ldots i_a} x^{i_1}_1 \ldots x^{i_a}_a \xi^1_{j_1} \ldots \xi^b_{j_b}$$

of a vector and b covector arguments and is as a rule identified with that functional.

With respect to the naturally defined operations of addition and multiplication by numbers (a, b)-tensors form a vector space $\mathbf{T}^b_a(\mathscr{V})$.

For any tensor S or R, of type (a, b) or (c, d) respectively, the formula

$$(S \otimes R)^{j_1 \dots j_{b+d}}_{i_1 \dots i_{a+c}} = S^{j_1 \dots j_b}_{i_1 \dots i_a} R^{j_{b+1} \dots j_{b+d}}_{i_{a+1} \dots i_{a+c}}$$

defines a tensor $S \otimes R$ of type $(a + c, b + d)$ called a *tensor product* of tensors S and R. This multiplication is associative and distributive under addition.

Moreover, a special operation on tensors is called *contraction* (see Lecture II.5).

Every vector can be naturally interpreted as a tensor of type (0, 1) and every covector as a tensor of type (1,0). For any basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of $\mathscr{V}$ and any indices $i_1, \dots, i_a, j_1, \dots, j_b = 1, \dots, n$, in the space $\mathrm{T}^b_a(\mathscr{V})$ a tensor

$$\mathbf{e}^{i_1} \otimes \dots \otimes \mathbf{e}^{i_a} \otimes \mathbf{e}_{j_1} \otimes \dots \otimes \mathbf{e}_{j_b},$$

is defined where $\mathbf{e}^{i_1}, \dots, \mathbf{e}^{i_a}$ are vectors of the conjugate basis for the space $\mathscr{V}^*$.

Tensors of this form make up a basis for a space $\mathrm{T}^b_a(\mathscr{V})$, the components of the tensor being the coordinates of a tensor in that basis, i.e.

$$S = S^{j_1 \dots j_b}_{i_1 \dots i_a} \mathbf{e}^{i_1} \otimes \dots \otimes \mathbf{e}^{i_a} \otimes \mathbf{e}_{j_1} \otimes \dots \otimes \mathbf{e}_{j_b}$$

for any tensor S.

We shall apply these general notions of linear algebra to the case where $\mathscr{V}$ is a tangent space $\mathbf{T}_p\mathscr{X}$ of a smooth manifold $\mathscr{X}$ at its point p.

Let $(U, h) = (U, x^1, \dots, x^n)$ be an arbitrary chart of $\mathscr{X}$ containing p. That chart defines a basis

$$\left(\frac{\partial}{\partial x^1}\right)_p, \dots, \left(\frac{\partial}{\partial x^n}\right)_p \tag{1}$$

in $\mathbf{T}_p\mathscr{X}$ and the conjugate basis

$$(dx^1)_p, \dots, (dx^n)_p$$

in the conjugate space $\mathbf{T}_p^*\mathscr{X}$. For every tensor S_p of type (a, b) on $\mathbf{T}_p\mathscr{X}$ we have a representation of the form

$$(2)\qquad S_p = S_{i_1\ldots i_a}^{j_1\ldots j_b}(dx^{i_1})_p \otimes \ldots \otimes (dx^{i_a})_p \otimes \left(\frac{\partial}{\partial x^{j_1}}\right)_p \otimes \ldots \otimes \left(\frac{\partial}{\partial x^{j_b}}\right)_p,$$

whose coefficients $S_{i_1\ldots i_a}^{j_1\ldots j_b}$ (i.e. the components of the tensor S_p in basis (1)) are called the *components of the tensor* S_p *in a chart* (U, h). (For typographical reasons we omit the index p in the notation of these components.)

Any other chart $(U', h') = (U', x^{1'}, \ldots, x^{n'})$ (with $p \in U'$) defines a basis

$$(3)\qquad \left(\frac{\partial}{\partial x^{1'}}\right)_p, \ldots, \left(\frac{\partial}{\partial x^{n'}}\right)_p.$$

for $\mathbf{T}_p\mathscr{X}$ connected with basis (1) by a transition matrix

$$(4)\qquad \left\|\left(\frac{\partial x^{i'}}{\partial x^{i}}\right)_p\right\|, \quad i, i' = 1, \ldots, n.$$

The components of S_p in charts (U, h) and (U', h') are related as follows:

$$(5)\qquad S_{i'_1\ldots i'_a}^{j'_1\ldots j'_b} = \left(\frac{\partial x^{i_1}}{\partial x^{i'_1}}\right)_p \ldots \left(\frac{\partial x^{i_a}}{\partial x^{i'_a}}\right)_p \left(\frac{\partial x^{j'_1}}{\partial x^{j_1}}\right)_p \ldots \left(\frac{\partial x^{j'_b}}{\partial x^{j_b}}\right)_p S_{i_1\ldots i_a}^{j_1\ldots j_b}.$$

If a tensor S_p is now given for any point $p \in \mathscr{X}$, then in (2) the components $S_{i_1\ldots i_a}^{j_1\ldots j_b}$ are functions of p. If these functions are smooth, i.e. expressed smoothly in the chart (U, h) in terms of the coordinates $x^1, \ldots, x^n$, then the correspondence $p \mapsto S_p$ is said to be a *(smooth) tensor field* (or in shorter form *tensor*) *of type* (a, b) *on a manifold* $\mathscr{X}$ Relation (5) for functions $S_{i_1\ldots i_a}^{j_1\ldots j_b}$ on $U \cap U'$ is of the form

$$(6)\qquad S_{i'_1\ldots i'_a}^{j'_1\ldots j'_b} = \frac{\partial x^{i_1}}{\partial x^{i'_1}} \ldots \frac{\partial x^{i_a}}{\partial x^{i'_a}} \frac{\partial x^{j'_1}}{\partial x^{j_1}} \ldots \frac{\partial x^{j'_b}}{\partial x^{j_b}} S_{i_1\ldots i_a}^{j_1\ldots j_b}$$

from which it follows that the condition that a tensor field should be smooth is independent of the choice of the chart.

Remark 1. For manifolds of a finite smoothness class C^r, $r \geqslant 1$, we run into problems such that the elements of matrix (4) are only functions of class C^{r-1}. Therefore the smoothness of tensor fields has to be understood only in the sense of C^{r-1}. To avoid these reservations we agreed in Lecture 6 to restrict our discussion to C^∞- and C^ω-manifolds for which difficulties of this sort do not arise.

For any open covering $\{U_\alpha\}$ of a manifold $\mathscr{X}$ every tensor field S defines a family of fields

$$S_\alpha = S \mid_{U_\alpha}$$

that have the property that for any indices α and β

$$S_\alpha = S_\beta \quad \text{on} \quad U_\alpha \cap U_\beta. \tag{7}$$

Conversely, given the fields S_α on U_α satisfying relations (7) (such fields are said to be *compatible on intersections*), the formula

$$S_p = (S_\alpha)_p \text{ if } p \in U_\alpha$$

correctly defines on $\mathscr{X}$ a tensor field S having the property that

$$S \mid_{U_\alpha} = S_\alpha$$

for any α (and hence smooth). We shall say that fields S_α *make up* a field S.

Remark 2. A tensor field on a manifold can be regarded as a correspondence associating every chart (U, h) of a manifold $\mathscr{X}$ with a collection of smooth functions $S^{j_1 \ldots j_b}_{i_1 \ldots i_a}$ on U and having the property that for any two charts (U, h) and (U', h') on the intersection $U \cap U'$ we have relation (6). This can be taken to be a definition of a tensor field. The advantage of this definition consists in that it can be formulated right after the introduction of the notion of a smooth manifold without any intermediate definitions, but its disadvantage lies in the absence of a direct formal connection (replaced by analogy) with the notion of a tensor in a vector space.

All algebraic operations on tensors (the operation of contraction among them) can be trivially extended to tensor fields. For example, a tensor product $S \otimes R$ of two tensor fields S and R is defined as follows:

$$(8) \qquad (S \otimes R)_p = S_p \otimes R_p.$$

Clearly, smooth fields always produce smooth fields.

In particular, we see that the *collection* $\mathsf{T}_a^b\mathscr{X}$ *of all tensor fields of type* (a, b) *on a manifold* $\mathscr{X}$ *is a vector space.*

This space is infinite dimensional (for $n > 0$).

For $(a, b) = (0, 0)$ tensor fields are nothing but smooth functions on $\mathscr{X}$ and the vector space $\mathsf{T}_0^0\mathscr{X}$ is a vector space $\mathsf{F}\mathscr{X}$ of smooth functions on $\mathscr{X}$. The vector space $\mathsf{F}\mathscr{X}$ is an algebra with respect to multiplication of functions, with the formula

$$(fS)_p = f(p) S_p, \quad f \in \mathsf{F}\mathscr{X}, \qquad S \in \mathsf{T}_a^b\mathscr{X}$$

(which is a special case of (8)) defining the operation of multiplication

$$\mathsf{F}\mathscr{X} \times \mathsf{T}_a^b\mathscr{X} \to \mathsf{T}_a^b\mathscr{X}$$

with respect to which, as a trivial check shows, the *vector space* $\mathsf{T}_a^b\mathscr{X}$ *is a module over an algebra* $\mathsf{F}\mathscr{X}$.

When $(a, b) = (0, 1)$ tensor fields are said to be *vector fields*. An example of a vector field on a coordinate neighbourhood U (considered as a manifold) is the field

$$(9) \qquad \frac{\partial}{\partial x^i}: p \mapsto \left(\frac{\partial}{\partial x^i}\right)_p, \quad i = 1, \ldots, n.$$

It is called the *i*th *coordinate vector field* on U.

When $(a, b) = (1, 0)$ tensor fields are said to be *covector fields*. An example of covector fields is the *i*th *coordinate covector field*

$$(10) \qquad dx^i: p \mapsto (dx^i)_p$$

on a coordinate neighbourhood U.

Formula (2) states that every tensor field S on U is uniquely decomposed into tensor products of vector and covector coordinate fields:

$$(11)\ S = S_{i_1 \ldots i_a}^{j_1 \ldots j_b}\, dx^{i_1} \otimes \ldots \otimes dx^{i_a} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_b}} \text{ on } U.$$

In particular, every vector field X on U is of the form

(12) $$X = X^i \frac{\partial}{\partial x^i},$$

and every covector field α is of the form

(13) $$\alpha = \alpha_i dx^i,$$

where X^i and α_i, $i = 1, \ldots, n$, are some smooth functions on U. (By tradition small Greek letters are used to denote covector fields while capital Latin letters from the end of the alphabet are used to denote vector fields.)

By definition the existence of decomposition (11) implies that *for any coordinate neighbourhood U the vector field $\mathsf{T}_a^b U$ is a free module over an algebra $\mathsf{F}U$ with basis*

$$dx^{i_1} \otimes \ldots \otimes dx^{i_a} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_b}}.$$

However, for arbitrary manifolds $\mathscr{X}$ the module $\mathsf{T}_b^a \mathscr{X}$ is not generally a free module (over an algebra $\mathsf{F}\mathscr{X}$), and its algebraic structure may be quite complicated.

Manifolds $\mathscr{X}$ for which all modules $\mathsf{T}_a^b \mathscr{X}$ are free are called *parallelizable.*

Let us consider vector fields more closely.

As was already stated in Lecture 12, every vector $A \in \mathbf{T}_p \mathscr{X}$ associates an arbitrary function f (defined and smooth in a neighbourhood of a point p) with some number Af, the derivative of that function with respect to a vector A. Hence it follows that for any vector field X on a manifold $\mathscr{X}$ and an arbitrary function $f \in \mathsf{F}\mathscr{X}$ the formula

(14) $$(Xf)(p) = X_p f, \quad p \in \mathscr{X}$$

defines on $\mathscr{X}$ some function Xf. It follows from the formulas of Lecture 12 for Af that in an arbitrary chart $(U, h) = (U, x^1, \ldots, x^n)$ of $\mathscr{X}$ the restriction of the function Xf on U is defined by

(15) $$Xf = X^i \frac{\partial f}{\partial x^i} \quad \text{on } U,$$

where X^i, $i = 1, \ldots, n$, are the components of a vector field X in the chart (U, h). The function Xf is therefore smooth on U and hence, U being arbitrary, on the whole of $\mathscr{X}$.

Thus formula (14) defines some (obviously linear) mapping X of an algebra $\mathsf{F}\mathscr{X}$ of smooth functions on a manifold $\mathscr{X}$ into itself. This mapping is called a *linear differential operator of the first order* on $\mathscr{X}$ generated by the vector field X. [This terminology is motivated by formula (15) whose comparison with formula (12) also explains the choice of the notation $\frac{\partial}{\partial x^i}$ for coordinate vector fields.]

Let $\mathscr{A}$ be an arbitrary algebra (not necessarily finite dimensional and associative).

Definition 1. A linear mapping

$$D\colon \mathscr{A} \to \mathscr{A}$$

of an algebra $\mathscr{A}$ into itself is said to be a *differentiation* if

$$D(ab) = Da \cdot b + a \cdot Db$$

for any elements $a, b \in \mathscr{A}$.

In particular, differentiations of an algebra $\mathsf{F}\mathscr{X}$ (usually called simply *differentiations on* $\mathscr{X}$) are linear mappings

$$D\colon \mathsf{F}\mathscr{X} \to \mathsf{F}\mathscr{X}$$

such that

$$D(fg) = Df \cdot g + f \cdot Dg \tag{16}$$

for any two smooth functions f and g on $\mathscr{X}$.

It is easy to see that a *linear differential operator* X: $\mathsf{F}\mathscr{X} \to \mathsf{F}\mathscr{X}$ *generated by a vector field* X *is a differentiation on* $\mathscr{X}$. Indeed, it follows directly from the rule for the differentiation of a product and from formula (15) that for any functions $f, g \in \mathsf{F}\mathscr{X}$ identity (16) holds on every coordinate neighbourhood U. It holds therefore on the entire manifold $\mathscr{X}$. □

It turns out that if a manifold $\mathscr{X}$ is Hausdorff, this exhausts all differentiations on $\mathscr{X}$.

Theorem 1. *Every differentiation D on a Hausdorff smooth manifold $\mathscr{X}$ of class C^∞ is generated by a vector field. This field is unique.*

To prove the theorem we need the following lemma.

Lemma 1. *Let $\mathscr{X}$ be a Hausdorff smooth manifold, U its open submanifold, and f a smooth function on U. For any point $p_0 \in U$ then there is a smooth function f_1 and a neighbourhood W of p_0 on $\mathscr{X}$ such that $\overline{W} \subset U$ and*

$$f = f_1 \text{ on } W.$$

It may also be assumed that $f_1 = 0$ outside U.

Proof. According to Proposition 2 of Lecture 14 there are open sets V and W in $\mathscr{X}$ such that

$$p_0 \in W, \quad \overline{W} \subset V, \overline{V} \subset U,$$

and there is a Urysohn function φ for a pair (V, W). For any point $p \in \mathscr{X}$ put

$$f_1(p) = \begin{cases} \varphi(p) f(p) & \text{if} \quad p \in U, \\ 0 & \text{if} \quad p \notin U. \end{cases}$$

Clearly, f_1 is smooth and coincides with the function f on W. Moreover, by construction $f_1 = 0$ outside U. □

Corollary 1. *For every neighbourhood U of an arbitrary point p_0 of a smooth Hausdorff manifold $\mathscr{X}$ there is a neighbourhood W of p_0 and a function φ smooth on $\mathscr{X}$ such that*

$$\varphi(p) = \begin{cases} 1 & \textit{if} \quad p \in W, \\ 0 & \textit{if} \quad p \notin U. \end{cases} \quad \square$$

Notice that $\varphi(p_0) = 1$ and $\overline{W} \subset U$.

Now let D be an arbitrary differentiation on a manifold $\mathscr{X}$.

We shall say that functions f and g smooth on $\mathscr{X}$ *are equal near a point* $p_0 \in \mathscr{X}$ if they assume equal values in some neighbourhood of that point.

Corollary 2. *If functions f and g are equal near a point of $\mathscr{X}$, then functions Df and Dg are also equal near that point.*

Proof. Let $f = g$ on a neighbourhood U of a point p_0. Consider a function $(f - g)\varphi$, where φ is the function on $\mathscr{X}$ of Corollary 1. Clearly $(f - g)\varphi = 0$ on the entire $\mathscr{X}$, i.e. the function $(f - g)\varphi$ is a zero of a vector space $\mathsf{F}\mathscr{X}$. Since the mapping D is linear, the function $D[(f - g)\varphi]$ is also zero therefore. Because

$$D[(f - g)\varphi] = (Df - Dg)\varphi + (f - g)D\varphi,$$

this proves that

$$(Df - Dg)\,\varphi + (f - g)\,D\varphi = 0$$

on the whole of $\mathscr{X}$ and in particular on the neighbourhood W of the point p_0 specified by Corollary 1. But $f = g$ and $\varphi = 1$ on W. Therefore $Df - Dg = 0$, and hence $Df = Dg$ on W.□

Remark 3. It also follows from Lemma 1 that *for any point* $p_0 \in U$ *and any tensor field* S *on* U *there exists a neighbourhood* $W \subset U$ *of that point and a tensor field* S_1 *on the entire* $\mathscr{X}$ *such that*

$$S = S_1 \text{ on } W.$$

To prove this it suffices to apply Lemma 1 to each component of S (and take the intersection of the corresponding neighbourhoods W).

The property expressed by Corollary 2 implies that the mapping D is **local**. It yields the following important proposition.

Proposition 1. *For any open set* $U \subset \mathscr{X}$ *there is a unique differentiation*

$$D_U\colon \mathsf{F}U \to \mathsf{F}U,$$

compatible with a differentiation D, *i.e. such that*

$$(17) \qquad D_U\,(f\mid_U) = Df\mid_U$$

for every function f *smooth on* $\mathscr{X}$.

Geometrically the latter property means that in the diagram

$$\begin{tikzcd}
\mathsf{F}\mathscr{X} \arrow[r] \arrow[d, "D"'] & \mathsf{F}U \arrow[d, "D_U"] \\
\mathsf{F}\mathscr{X} \arrow[r] & \mathsf{F}U
\end{tikzcd}$$

whose horizontal arrows are restriction mappings, the same mapping $\mathsf{F}\mathscr{X} \to \mathsf{F}U$ results when moving in the two possible paths — from the top left to the bottom right corner. Diagrams having this property are called *commutative* (we have already encountered commutative diagrams in Lecture 3).

Proof of Proposition 1. Let the differentiation D_U exist and let g be an arbitrary function smooth on U. By Lemma 1 there is a function $g_1 = g_{1,p_0}$ for any point $p_0 \in U$ on $\mathscr{X}$ that coincides with a function g near p_0. According to property (17) $D_U g = Dg_1 \mid_U$ and hence, in particular,

$$(D_U g)(p_0) = (Dg_1)(p_0). \tag{18}$$

On the other hand, if g_1^* is another function smooth on $\mathscr{X}$ which coincides near p_0 with g (and hence with g_1), then, D being a local mapping, we have near p_0 an equation $Dg_1 = Dg_1^*$ and hence, in particular, an equation

$$(Dg_1)(p_0) = (Dg_1^*)(p_0). \tag{19}$$

Thus the right-hand side of (17) is independent of the choice of function g_1 and is defined exclusively by the function g and the point p_0. This means that the function $D_U g$ on U depends only on g. Consequently, the mapping D_U: $g \mapsto D_U g$ is unique.

To prove its existence, we assume formula (18) to be a definition of the function $D_U g$. According to equation (19) this definition is correct. Moreover, if $g_1 = g$ on a neighbourhood W of p_0, then for any point $p \in W$ we can use the function $g_1 = g_{1,p_0}$ as function $g_{1,p}$ to calculate values $(D_U g)(p)$ of $D_U g$ at p. This means that $D_U g = Dg_1$ not only at p_0 but also near this point. Thus the function $D_U g$ coincides with some function smooth on $\mathscr{X}$ near each point of U. The function $D_U g$ is therefore smooth on U, i.e. the correspondence $g \mapsto D_U g$ is a mapping

$$D_U\colon \mathsf{F}U \to \mathsf{F}U.$$

If f and g are two functions smooth on U and f_1 and g_1 are functions smooth on $\mathscr{X}$ that coincide near p_0 with f and g, respectively, then the function $f_1 + g_1$ coincides near p_0 with $f + g$. Therefore

$$\begin{aligned}[D_U (fg)](p_0) &= [D(f_1 g_1)](p_0)\\ &= [Df_1 \cdot g_1 + f_1 \cdot Dg_1](p_0)\\ &= [D_U f \cdot g + f \cdot D_g](p_0),\end{aligned}$$

and hence the mapping D_U, which is obviously linear, is a differentiation on U.

Finally, if f is a function smooth on $\mathscr{X}$, then for the function $g = f\,|_U$ the role of g_1 (for any point $p_0 \in U$) can be played by the function f. Therefore

$$D_U (f\,|_U) = Df\,|\,_U. \square$$

As a rule we shall write simply Dg rather than $D_U g$.

To prove Theorem 1 we shall also need the following lemma.

Lemma 2. *Let $(U, h) = (U, x^1, \ldots, x^n)$ be an arbitrary chart of a C^∞-manifold $\mathscr{X}$. Then near any point $p_0 \in U$ every function f smooth (on U) allows a representation of the form*

$$f = f(p_0) + (x^i - x_0^i)\, f_i, \tag{20}$$

where $f_1, \ldots, f_n$ are some functions smooth (near p_0) and $x_0^1, \ldots, x_0^n$ are the coordinates of p_0.

Proof. Let $\mathbf{x}_0 = h(p)$, and let $r > 0$ be a number such that any point $\mathbf{x} \in \mathbb{R}^n$ for which $|\mathbf{x} - \mathbf{x}_0| < r$ belongs to an open set $h(U)$. Then for the value $f(\mathbf{x})$ at each such point $\mathbf{x}$ of an arbitrary function f smooth in $h(U)$ the formula

$$f(\mathbf{x}) - f(\mathbf{x}_0) = \int_0^1 \frac{\partial}{\partial s} f(\mathbf{x}_0 + s(\mathbf{x} - \mathbf{x}_0))\, ds$$

$$= (x^i - x_0^i) \int_0^1 \frac{\partial f}{\partial x^i} (\mathbf{x}_0 + s(\mathbf{x} - \mathbf{x}_0))\, ds,$$

holds, i.e.

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (x^i - x_0^i)\, f_i(\mathbf{x}),$$

where

$$f_i(\mathbf{x}) = \int_0^1 \frac{\partial f}{\partial x^i} (\mathbf{x}_0 + s(\mathbf{x} - \mathbf{x}_0))\, ds.$$

To complete the proof it remains to use a diffeomorphism h to change from functions on $h(U)$ to functions on U. $\square$

For every differentiation D of an arbitrary algebra $\mathscr{A}$ with identity element 1 the equality

$$D1 = D(1 \cdot 1) = D1 \cdot 1 + 1 \cdot D1 = D1 + D1 = 2D1$$

holds and hence

$$D1 = 0.$$

Since D is a linear mapping, it follows that $Da = 0$ *for every element* $a \in \mathbb{R}$ *of the ground field.*

For differentiations on a smooth manifold $\mathscr{X}$ (i.e. differentiations of an algebra $\mathsf{F}\mathscr{X}$) this means that *every differentiation on* $\mathscr{X}$ *makes any constant function vanish.*

Now we can prove Theorem 1.

Proof of Theorem 1. As ever we first prove the uniqueness.

Let a differentiation D on a manifold $\mathscr{X}$ be generated by a vector field X. This means that in every chart $(U, x^1, \ldots, x^n)$ of $\mathscr{X}$ for any function f smooth on U the equality

$$Df = X^i \frac{\partial f}{\partial x^i} \quad \text{on } U,$$

holds where X^i, $i = 1, \ldots, n$, are the components of a vector field X on U (where D in fact designates D_U). In particular,

$$Dx^i = X^i, \ i = 1, \ldots, n,$$

which proves that the field X is unique.

Now let D be an arbitrary differentiation on a manifold $\mathscr{X}$ (assumed to be Hausdorff and of class C^∞). For any chart $(U, x^1, \ldots, x^n)$ consider on U smooth functions

$$X^i = Dx^i, \ i = 1, \ldots, n.$$

Let f be an arbitrary function smooth on U. Applying to its representation (20) an operator D (or more exactly an operator D_W, where W is a neighbourhood of a point p_0 in which (19) holds) we get

$$Df = X^i f_i + (x^i - x_0^i)\, Df_i \quad \text{on } W$$

and hence

$$(Df)(p_0) = X^i(p_0)\, f_i(p_0). \tag{21}$$

(We have the right to write D instead of D_W because the operator D is local.)

Equation (21) is true not only for differentiations on $\mathscr{X}$ but also for any differentiation on U. In particular it is true for the differentiation $\frac{\partial}{\partial x^i}$. Therefore

$$\left(\frac{\partial}{\partial x^i} f\right)(p_0) = f_i(p_0),$$

and hence, by virtue of formula (21),

$$(Df)(p_0) = X^i(p_0)\frac{\partial f}{\partial x^i}(p_0) = \left(X^i \frac{\partial f}{\partial x^i}\right)(p_0).$$

Since p_0 is an arbitrary point on a neighbourhood U, this proves that

$$Df = X^i \frac{\partial f}{\partial x^i} \text{ on } U,$$

i.e. the differentiation D is generated in U by a vector field X_U with components $X^1, \ldots, X^n$.

Similarly, for any other chart $(U', x^{1'}, \ldots, x^{n'})$ a differentiation D on U' is generated by a vector field $X_{U'}$ with components $X^{i'} = Dx^{i'}$. The operator D being local, for any function f on the intersection $U \cap U'$ the equality

$$X^{i'} \frac{\partial f}{\partial x^{i'}} = X^i \frac{\partial f}{\partial x^i}$$

holds. Applying this equality to transition functions $x^{i'} = x^{i'}(x^1, \ldots, x^n)$ we immediately get

$$X^{i'} = X^i \frac{\partial x^{i'}}{\partial x^i},$$

i.e. on $U \cap U'$ vector fields X_U and $X_{U'}$ coincide. Putting for any point p of a manifold $\mathscr{X}$

$$X_p = (X_U)_p,$$

where U is an arbitrary coordinate neighbourhood of p, we therefore correctly define on $\mathscr{X}$ a vector field X having the property that $X|_U = X_U$ for any coordinate neighbourhood U. Hence this field is, first, smooth and, second, generates a differentiation D. □

Test question. Where in this proof is the assumption that $\mathscr{X}$ is a C^∞-manifold used?

On the basis of Theorem 1 vector fields on $\mathscr{X}$ are usually identified with differentiations. (Which, in particular, justifies a posteriori the use of Xf to denote the result of applying the differentiation generated by a vector field X to the function f.)

It is clear that for any algebra $\mathcal{A}$ the sum of two differentiations and the product of a differentiation by a number are also differentiations, i.e. *the set* Der $\mathcal{A}$ *of all differentiations of an algebra* $\mathcal{A}$ *is a vector space* (a subspace of the vector space $\operatorname{End}_{\text{vec}}\mathcal{A}$ of all linear operators $\mathcal{A} \to \mathcal{A}$).

For any two linear operators $D_1, D_2\colon \mathcal{A} \to \mathcal{A}$ the operator

$$[D_1, D_2] = D_1D_2 - D_2D_1$$

is called their *commutator* (or *Lie bracket*).

It is easy to see that *for every algebra* $\mathcal{A}$ *the commutator* $[D_1, D_2]$ *of any two differentiations* $D_1, D_2\colon \mathcal{A} \to \mathcal{A}$ *is also a differentiation.* Indeed,

$$\begin{aligned}[D_1, D_2](ab) &= D_1(D_2(ab)) - D_2(D_1(ab)) \\ &= D_1(D_2a\cdot b + a\cdot D_2b) - D_2(D_1a \\ &\quad \times b + a\cdot D_1b) \\ &= D_1D_2a\cdot b + D_2a\cdot D_1b + D_1a\cdot D_2b \\ &\quad + a\cdot D_1D_2b \\ &\quad - D_2D_1a\cdot b - D_1a\cdot D_2b - D_2a \\ &\quad \times D_1b - a\cdot D_2D_1b \\ &= (D_1D_2 - D_2D_1)\, a\cdot b + a\cdot (D_1D_2 \\ &\quad - D_2D_1)\, b \\ &= [D_1, D_2]\, a\cdot b + a\cdot [D_1, D_2]\, b\end{aligned}$$

for arbitrary elements $a, b \in \mathcal{A}$.☐

Since the commutation operation $D_1, D_2 \mapsto [D_1, D_2]$ is obviously linear in D_1 and D_2, this means that *with respect to the commutation operation the vector space* Der $\mathcal{A}$ *of all differentiations of an algebra* $\mathcal{A}$ *is itself an algebra.*

The *commutation operation is anticommutative*, i.e.

$$[D_1, D_2] = -[D_2, D_1]$$

for any two linear operators D_1 and D_2 (which are even not differentiations).

Moreover, *for any three operators* D_1, D_2 *and* D_3 *we have the identity*

(22) $$[[D_1, D_2], D_3] + [[D_2, D_3], D_1] + [[D_3, D_1], D_2] = 0.$$

Indeed,

$$\begin{aligned}[[D_1, D_2], D_3] &= (D_1D_2 - D_2D_1)\, D_3 - D_3\,(D_1D_2 - D_2D_1) \\ &= D_1D_2D_3 - D_2D_1D_3 - D_3D_1D_2 + D_3D_2D_1,\end{aligned}$$

and hence the left-hand side of (22) is the sum of twelve operators of the form $D_iD_jD_k$ in which each of the operators occurs twice with opposite signs. This sum is therefore zero. □

Identity (22) is called a *Jacobi identity*.

All this motivated the following general definition.

Definition 2. An algebra in which multiplication is anticommutative and satisfies the Jacobi identity is called a *Lie algebra*.

Thus we have proved that the *algebra* Der $\mathcal{A}$ *is a Lie algebra*.

Since differentiations of an algebra of smooth functions $\mathsf{F}\mathscr{X}$ due to identifications are nothing but vector fields on $\mathscr{X}$, we see in particular that *vector fields on a Hausdorff C^∞-manifold $\mathscr{X}$ form a Lie algebra.*

This Lie algebra is usually designated $\mathfrak{a}\mathscr{X}$.

By definition

$$[X, Y]\, f = X\,(Yf) - Y\,(Xf)$$

for any vector fields X, Y on $\mathscr{X}$ and any function $f \in \mathsf{F}\mathscr{X}$.

It follows that

(23) $$[gX, Y] = g\,[X, Y] - Yg\cdot X$$

for any function $g \in \mathsf{F}\mathscr{X}$. Indeed,

$$\begin{aligned}[gX, Y]\, f &= gX\,(Yf) - Y\,(gXf) \\ &= gX\,(Yf) - Yg\cdot Xf - gY\,(Xf) \\ &= g\,[X, Y]\, f - Yg\cdot Xf\end{aligned}$$

for every function $f \in \mathsf{F}\mathscr{X}$. □

If

$$X = X^i \frac{\partial}{\partial x^i} \quad \text{and} \quad Y = Y^j \frac{\partial}{\partial x^j},$$

in a chart $(U, x^1, \ldots, x^n)$, then for any function f smooth on U

$$\begin{aligned}[X, Y] f = X(Yf) - Y(Xf) &= X\left(Y^j \frac{\partial f}{\partial x^j}\right) - Y\left(X^i \frac{\partial f}{\partial x^i}\right) \\ &= X^i \frac{\partial}{\partial x^i}\left(Y^j \frac{\partial f}{\partial x^j}\right) - Y^j \frac{\partial}{\partial x^j}\left(X^i \frac{\partial f}{\partial x^i}\right) \\ &= X^i \frac{\partial Y^j}{\partial x^i} \frac{\partial f}{\partial x^j} + X^i Y^j \frac{\partial^2 f}{\partial x^i \partial x^j} - Y^j \frac{\partial X^i}{\partial x^j} \frac{\partial f}{\partial x^i} \\ &\quad - Y^j X^i \frac{\partial^2 f}{\partial x^i \partial x^j} = \left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}\right) \frac{\partial f}{\partial x^i}.\end{aligned}$$

This proves that in *every chart* $(U, x^1, \ldots, x^n)$ *the components* $[X, Y]^i$ *of a vector field* $[X, Y]$ *can be expressed by the formula*:

$$(24) \qquad [X, Y]^i = X^j \frac{\partial Y^i}{\partial x^j} - X^j \frac{\partial X^i}{\partial x^j}, \quad i = 1, \ldots, n,$$

where $X^1, \ldots, X^n, Y^1, \ldots, Y^n$ *are the components of the fields* X *and* Y, *respectively.*

Remark 4. We have already used the term "Lie algebra" in Lecture 6 in connection with vector spaces of matrices specified by the definition of matrix Lie groups (see Definition 1 of Lecture 11). With respect to the commutation operation $[B_1, B_2] = B_1B_2 - B_2B_1$ of matrices the vector space $\mathrm{Mat}_n(\mathbb{R}) = \mathfrak{gl}_n(\mathbb{R})$ of all matrices of order n is a Lie algebra. A Lie algebra is therefore any of its subalgebras (a linear subspace $\mathfrak{g}$ closed under this operation) as well. On the other hand, if $\mathfrak{g}$ is a subspace of Definition 1, Lecture 11, and if $B_1, B_2 \in \mathfrak{g}$, then by Condition 1° of that definition matrices e^{tB_1} and e^{tB_2} for any $t \in \mathbb{R}$ belong to a group $\mathscr{G}$. So to this group also belongs the matrix

$$\begin{aligned}A_t &= e^{tB_1} e^{tB_2} e^{-tB_1} e^{-tB_2} \\ &= (E + B_1 t + \ldots)(E + B_2 t + \ldots)(E - B_1 t \\ &\quad + \ldots)(E - B_2 t + \ldots) = E + [B_1, B_2] t^2 + \ldots,\end{aligned}$$

where the dots designate terms having a power $\geqslant 3$ in t. Since $A_t \to E$ as $t \to 0$, the norm of the matrix $A_t - E$ is less than 1, with t sufficiently small, and hence a matrix

$$B_t = \ln A_t = (A_t - E) - \frac{(A-E)^2}{2} + \ldots = [B_1, B_2]\, t^2 + \ldots$$

is defined.

Since $B_t \to 0$ as $t \to 0$, with t sufficiently small, for the matrix norm $|B_t|$ we have $|B_t| < \ln 2$, and hence, by Condition 2° of Definition 1, Lecture 11, the matrix B_t belongs to a subspace $\mathfrak{g}$. But then to this subspace belongs the matrix $\frac{B_t}{t^2}$ and therefore the matrix

$$\lim_{t \to 0} \frac{B_t}{t^2} = \lim_{t \to 0} ([B_1, B_2] + \ldots) = [B_1, B_2].$$

As well $\mathfrak{g}$ is a Lie algebra (a subalgebra of the Lie algebra $\mathfrak{gl}_n(\mathbb{R})$). Subalgebras of $\mathfrak{gl}_n(\mathbb{R})$ are called *matrix Lie algebras.*

Remark 5. It should be kept in mind that there exist matrix Lie algebras that are not Lie algebras of any matrix Lie group. We shall return to this question in the next semester.

Lecture 17

Integral curves of vector fields · Vector fields and flows · Transfer of vector fields using diffeomorphisms · The Lie derivative of a tensor field

---

By a *curve* on a smooth manifold $\mathscr{X}$ we mean an arbitrary smooth mapping

$$\gamma\colon\ (a, b) \to \mathscr{X} \tag{1}$$

into $\mathscr{X}$ of some interval (a, b) of the axis $\mathbb{R}$ (see the discussion of the concept of a curve in Lecture 1).

The differential $(d\gamma)_t$ of γ at every point $t \in (a, b)$ is a linear mapping of a one-dimensional space $\mathbf{T}_t\,(a, b) = \mathbb{R}$ into a space $\mathbf{T}_{\gamma(t)}\mathscr{X}$ and is uniquely characterized by a vector

$$\dot{\gamma}\,(t) = (d\gamma)_t\left(\frac{\partial}{\partial t}\right)_t \tag{2}$$

of $\mathbf{T}_{\gamma(t)}\mathscr{X}$ into which it carries a basis vector $\left(\frac{\partial}{\partial t}\right)$ of $\mathbf{T}_t\,(a, b)$.

Definition 1. A vector $\dot{\gamma}\,(t)$ is called the *tangent vector* of a curve γ at a point t. (Somewhat loosely, the vector $\dot{\gamma}\,(t)$ is also called the tangent vector of a curve γ at a point $p = \gamma\,(t)$.)

Now let X be a vector field on a manifold $\mathscr{X}$. By definition every point $p \in \mathscr{X}$ is associated by X with some vector $X_p \in \mathbf{T}_p\mathscr{X}$.

Definition 2. A curve γ is said to be an *integral curve* (or *trajectory*) of a vector field X if

$$\dot{\gamma}\,(t) = X_{\gamma(t)} \quad \text{for any } t, \quad a < t < b. \tag{3}$$

A curve γ is said to be *contained* in a chart $(U, h) = (U, x^1, \ldots, x^n)$ if $\gamma(t) \in U$ for any t, $a < t < b$. Such a curve is given by n smooth functions

$$x^i = x^i(t), \quad a < t < b, \quad i = 1, \ldots, n, \tag{4}$$

and for its tangent vector $\dot{\gamma}(t)$ we have

$$\dot{\gamma}(t) = \dot{x}^i(t)\left(\frac{\partial}{\partial x^i}\right)_{\gamma(t)}$$

(prove it!). For a curve in U therefore equation (13) is equivalent to a system of n differential equations

$$\dot{x}^i(t) = X^i(x^1(t), \ldots, x^n(t)), \quad i = 1, \ldots, n, \tag{5}$$

of the first degree, where X^i, $i = 1, \ldots, n$, are the components of a vector field X in a chart (U, h) (or more exactly their expressions in terms of coordinates $x^1, \ldots, x^n$).

We thus see that equations of form (3) are a generalization to the case of arbitrary manifolds of the concept of a system of first-degree differential equations which is given in the domain of $\mathbb{R}^n$. They are called *differential equations on a manifold* $\mathscr{X}$. Their theory belongs to the theory of differential equations and is on the whole beyond the scope of the book.

Nevertheless, for the completeness of presentation and to demonstrate the specific character of general equations of form (3) we now state and prove the unique existence theorem for their solutions.

It is clear that for every subinterval $I' \subset I$ of the interval $I = (a, b)$ the restriction $\gamma \mid_{I'}$ of the integral curve γ to I' is also an integral curve of a vector field X. An integral curve (1) is said to be *maximal* if it is not a restriction to any integral curve defined on a larger interval.

Let $t_0 \in \mathbb{R}$. Curve (1) is said to *pass at* $t = t_0$ *through a point* $p \in \mathscr{X}$ if, first, this curve is defined on an interval (a, b) of the axis $\mathbb{R}$ such that $a < t_0 < b$ and, second, $\gamma(t_0) = p$.

Theorem 1. *If a manifold* $\mathscr{X}$ *is Hausdorff, then for any point* $p_0 \in \mathscr{X}$ *and any vector field* X *on* $\mathscr{X}$ *there is a unique* *maximal integral curve* γ: $I \to \mathscr{X}$ *passing at* $t = t_0$ *through* p.

Proof. Let Γ be the set of all possible integral curves γ of a vector field X passing at $t = t_0$ through a point p_0. Since in an arbitrary chart (U, h) containing p_0 equation (3) is equivalent to system (5) of ordinary differential equations, according to the unique existence theorem for solutions of ordinary differential equations there is at least one integral curve γ of X with $\gamma(t_0) = p_0$ in a neighbourhood U. This means that the *set* Γ *is not empty*.

Now let γ_1: $I_1 \to \mathscr{X}$ and γ_2: $I_2 \to \mathscr{X}$ be two arbitrary integral curves of Γ. Consider an intersection $I_0 = I_1 \cap I_2$ of intervals I_1 and I_2 which is obviously an interval of the axis $\mathbb{R}$ containing a point p_0 and its subset C consisting of all points $t \in I_0$ for which $\gamma_1(t) = \gamma_2(t)$. Under the hypothesis $t_0 \in C$, so that the *set* C *is not empty*.

Let $t_1 \in C$, and let (U, h) be an arbitrary chart containing a point $p_1 = \gamma_1(t_1) = \gamma_2(t_1)$. According to the unique existence theorem for solutions of ordinary differential equations there exists a unique integral curve γ: $I \to \mathscr{X}$ of X on some interval I of $\mathbb{R}$ containing t_1, for which $\gamma(t_1) = p_1$. We may assume without loss of generality that $I \subset I_0$. But then restrictions $\gamma_1|_I$ and $\gamma_2|_I$ of curves γ_1 and γ_2 on the interval I will be such a curve γ, and hence by virtue of the uniqueness they will coincide. By definition this means that $I \subset C$. Thus for any point $t_1 \in C$ there is an interval I containing that point such that $I \subset C$. Consequently, the *set* C *is open* (in $\mathbb{R}$ and hence in I_0 as well).

Now consider a mapping $\gamma_1 \times \gamma_2$: $I_0 \to \mathscr{X} \times \mathscr{X}$ defined as follows:

$$(\gamma_1 \times \gamma_2)(t) = (\gamma_1(t), \quad \gamma_2(t)), \quad t \in I.$$

Clearly, the mapping $\gamma_1 \times \gamma_2$ is continuous and the set C is nothing but the inverse image $(\gamma_1 \times \gamma_2)^{-1}\Delta$ of the diagonal $\Delta = \{(p, p), p \in \mathscr{X}\}$ of the product $\mathscr{X} \times \mathscr{X}$ under that mapping (see Lecture 15). Since for a Hausdorff manifold $\mathscr{X}$ the diagonal Δ is closed in $\mathscr{X}$, it follows that the *set* C *is closed in* I_0.

Since the interval I_0 is connected, this proves that $C = I_0$. This means that *any two integral curves of* Γ

coincide on the common interval of their domains of definition.

It follows that denoting by I_γ an interval on which a curve $\gamma \in \Gamma$ is defined and putting

$$I = \bigcup_{\gamma \in \Gamma} I_\gamma$$

and

$$\gamma_0(t) = \gamma(t) \text{ if } t \in I_\gamma$$

we correctly define a unique maximal integral curve $\gamma_0 \colon I \to \mathscr{X}$ of X for which $\gamma(t_0) = p_0$.□

Problem 1. Construct a two-dimensional non-Hausdorff manifold $\mathscr{X}$ and a vector field X on $\mathscr{X}$ such that for some point $p_0 \in \mathscr{X}$ there exist two distinct integral curves of X passing at $t = t_0$ through the point p_0.

Let $t_0 = 0$.

For an arbitrary point $p \in \mathscr{X}$ we denote by γ_p^X a maximal integral curve of X passing at $t = 0$ through the point p, and for any $t \in \mathbb{R}$ for which the point $\gamma_p^X(t)$ is defined we put

$$\varphi_t^X(p) = \gamma_p^X(t).$$

Thus φ_t^X is a mapping into a manifold $\mathscr{X}$ of a subset $D_t \subset \mathscr{X}$ consisting of points $p \in \mathscr{X}$ for which the point $\gamma_p^X(t)$ is defined.

Problem 2. Prove that the *set D_t is open (in $\mathscr{X}$) and (for $D_t \neq \varnothing$) the mapping $\varphi_t^X \colon D_t \to \mathscr{X}$ is smooth.* [**Hint.** Use the theorem on the dependence of solutions of systems of differential equations on the initial data, which is proved in the course in the theory of differential equations.]

Problem 3. Prove that mappings $\varphi_t = \varphi_t^X$ have the following properties:

(a) There exists a continuous function $\varepsilon \colon \mathscr{X} \to \mathbb{R}$ assuming positive values such that $p \in D_t$ provided $|t| < \varepsilon(p)$.

(b) The mapping φ_0 is defined on the entire $\mathscr{X}$ (i.e $D_0 = \mathscr{X}$) and is an identity mapping id of a manifold $\mathscr{X}$.

(c) If $\varphi_t(p) \in D_s$ (in particular, if $|s| < \varepsilon(\varphi_t(p))$, then $p \in D_{s+t}$ and

$$\varphi_s(\varphi_t(p)) = \varphi_{s+t}(p). \tag{6}$$

[**Hint.** To prove statement (a) use the theorem on the dependence of solutions of systems of differential equations on the initial data. Statement (b) is obvious, and to prove statement (c) it suffices to notice that both curves $s \mapsto \gamma^X_{\varphi_t(p)}(s)$ and $s \mapsto \gamma^X_p(s+t)$ are maximal integral curves of X passing for $s = 0$ through a point $\varphi_t(p)$.]

Somewhat loosely, property (c) is usually written as an identity

$$\varphi_s \circ \varphi_t = \varphi_{s+t}.$$

Definition 3. A family of smooth mappings $\varphi_t \colon D_t \to \mathscr{X}$ having properties (a), (b) and (c) is called a *flow* on a manifold $\mathscr{X}$.

We thus see that *every vector field* $X \in a\mathscr{X}$ *induces some flow* $\{\varphi^X_t\}$ *on* $\mathscr{X}$.

Conversely, every flow $\{\varphi_t\}$ defines by the formula

$$X_p = \dot{\gamma}_p(0), \quad p \in \mathscr{X},$$

where γ_p is a curve $t \mapsto \varphi_t(p)$, $|t| < \varepsilon(p)$, some vector field X on a manifold $\mathscr{X}$.

A flow $\{\varphi'_t \colon D'_t \to \mathscr{X}\}$ is said to be a *part* of a flow $\{\varphi_t \colon D_t \to \mathscr{X}\}$ if $D'_t \subset D_t$ and $\varphi_t \mid_{D'_t} = \varphi'_t$ for any $t \in \mathbb{R}$. A flow $\{\varphi_t\}$ is said to be *maximal* if it is not a part of any other flow.

It is clear that:

1° The flow $\{\varphi_t\}$ and any part of it $\{\varphi'_t\}$ generate the same vector field X.

2° The flow $\{\varphi^X_t\}$ induced by the vector field X is maximal.

Therefore the *formula*

$$\textit{field } X \Rightarrow \textit{flow } \{\varphi^X_t\}$$

establishes a one-to-one correspondence between vector fields and maximal flows on $\mathscr{X}$.

Since the function ε is continuous and hence

$$\lim_{t \to 0} \varepsilon(\varphi_t(p)) = \varepsilon(p) > 0$$

for any point $p \in \mathscr{X}$, there is a continuous function $\delta \colon \mathscr{X} \to \mathbb{R}$ such that

$$|t| < \varepsilon(\varphi_t(p)) \text{ for } |t| < \delta(p).$$

Let O_t be an open subset of a manifold $\mathscr{X}$ that consists of all points $p \in \mathscr{X}$ for which $|t| < \delta(p)$. Then there is $\delta_0 > 0$, namely, $\delta_0 = \delta(p_0)$ such that $p_0 \in O_t$ for $|t| < \delta_0$ (so that, for t sufficiently small, the set O_t is trivially not empty).

Since for $p \in O_t$, i.e. for $|t| < \delta(p)$, a point $\varphi_{-t}(\varphi_t(p))$ is defined which by property (c) coincides with a point p, a restriction of a mapping φ_t on a set O_t is a bijective mapping of that set onto an (obviously open) set $O'_t = \varphi_t O_t$ (with inverse mapping $\varphi_{-t}|_{O'_t}$). Since both mappings φ_t and φ_{-t} are smooth by definition, this proves that *for every* $t \in \mathbb{R}$ *(for which the set* O_t *is not empty) the mapping* φ_t *is a diffeomorphism* $O_t \to O'_t$.

Now let $\varphi\colon \mathscr{X} \to \mathscr{Y}$ be an arbitrary diffeomorphism of smooth manifolds, and let S be tensor field of type (a, b) on a manifold $\mathscr{Y}$. By definition in every chart (V, k) on $\mathscr{Y}$ the field S has components $S^{j_1 \ldots j_b}_{i_1 \ldots i_a}$ which are smooth functions on V. On the other hand, for any chart (U, h) on $\mathscr{X}$ a pair (V, k), where $V = \varphi U$, and $k = h \circ \varphi^{-1}$, is obviously a chart on $\mathscr{Y}$. Making use of this we define a tensor field $\varphi^* S$ on $\mathscr{X}$ assuming that in a chart (U, h) it has components

$$(\varphi^* S)^{j_1 \ldots j_b}_{i_1 \ldots i_a} = S^{j_1 \ldots j_b}_{i_1 \ldots i_a} \circ (\varphi|_U), \tag{7}$$

where $S^{j_1 \ldots j_b}_{i_1 \ldots i_a}$ are the components of the field S in a chart $(\varphi U, h \circ \varphi^{-1})$. Since for any two charts (U, h) and (U', h') on $\mathscr{X}$ a mapping of transition $h' \circ h^{-1}$ coincides with the mapping of transition $(h' \circ \varphi^{-1}) \circ (h \circ \varphi^{-1})^{-1}$, for charts $(\varphi U, h \circ \varphi^{-1})$ and $(\varphi U', h' \circ \varphi^{-1})$, the functions $(\varphi^* S)^{j_1 \ldots j_b}_{i_1 \ldots i_a}$ in distinct charts are related (at the intersection of these charts) by the same tensor transformation law as the components $S^{j_1 \ldots j_b}_{i_1 \ldots i_a}$ of the field S. Consequently, these functions are indeed the components of some tensor field $\varphi^* S$.

A field $\varphi^* S$ is said to be the result of a transfer of a field S from $\mathscr{Y}$ to $\mathscr{X}$ via a diffeomorphism φ.

Example 1. If a field S (a smooth function f on $\mathscr{Y}$) is of type (0, 0), then

$$\varphi^* f = f \circ \varphi. \tag{8}$$

Example 2. If a field S (a vector field X) is of type (0, 1) then

$$(\varphi^* X)_p = (d\varphi_p)^{-1} X_{\varphi(p)} \tag{9}$$

for any point $p \in \mathscr{X}$, where $(d\varphi_p)^{-1}$: $\mathbf{T}_{\varphi(p)}\mathscr{Y} \to \mathbf{T}_p\mathscr{X}$ is a mapping inverse to an isomorphism $d\varphi_p$: $\mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\varphi(p)}\mathscr{Y}$.

Example 3. If a field S (a covector field α), is of type (1, 0) then

$$(\varphi^* \alpha)_p = (d\varphi_p)^* \, \alpha_{\varphi(p)} \tag{10}$$

for any point $p \in \mathscr{X}$, where $(d\varphi_p)^*$ is a mapping $\mathbf{T}^*_{\varphi p}\mathscr{Y} \to \mathbf{T}^*_p\mathscr{X}$ conjugate to a mapping $d\varphi_p$: $\mathbf{T}_p\mathscr{X} \to \mathbf{T}_{\varphi(p)}\mathscr{Y}$.

Problem 4. Describe in a similar way tensors $(\varphi^* S)_p$ for a tensor field of arbitrary type (a, b). [**Hint.** The correspondence $S \mapsto \varphi^* S$ preserves all algebraic operations on tensor fields. In particular,

$$\varphi^* (S \otimes T) = \varphi^* S \otimes \varphi^* T \tag{11}$$

for any tensor fields S and T on $\mathscr{X}$.]

Remark 1. It can be easily seen that *for any vector fields* X, $Y \in \mathfrak{a}\mathscr{Y}$ *and any diffeomorphism* φ: $\mathscr{X} \to \mathscr{Y}$ *the formula*

$$[\varphi^* X, \varphi^* Y] = \varphi^* [X, Y],$$

holds, i.e. a mapping φ^*: $\mathfrak{a}\mathscr{Y} \to \mathfrak{a}\mathscr{X}$ (obviously linear) is an isomorphism of a Lie algebra $\mathfrak{a}\mathscr{Y}$ onto a Lie algebra $\mathfrak{a}\mathscr{X}$. (To prove this it suffices to notice that in local coordinates in which φ preserves the coordinates both sides of the formula obviously coincide.)

Remark 2. Note that, as can be seen directly from (9), a transfer of a vector field is only possible via a diffeomorphism. On the contrary formula (10) has meaning for any smooth mapping φ: $\mathscr{X} \to \mathscr{Y}$. Covector fields can therefore be transferred via arbitrary mappings. We shall return to this question in the next lecture.

Remark 3. Denoting a field X by Y and a field φ^*X by X we can write (9) as follows:

$$Y_{\varphi(p)} = (d\varphi)_p X_p. \tag{9'}$$

In this form it makes sense for any smooth mapping $\varphi\colon \mathscr{X} \to \mathscr{Y}$. If equality (9′) at every point $p \in \mathscr{X}$ holds for fields $X \in \mathfrak{a}\mathscr{X}$, $Y \in \mathfrak{a}\mathscr{Y}$, then the fields X and Y are said to be *φ-connected*.

Problem 5. Prove that *if fields X_1, $X_2 \in \mathfrak{a}\mathscr{X}$ are φ-connected with fields Y_1, $Y_2 \in \mathfrak{a}\mathscr{Y}$, respectively, then the field $[X_1, X_2]$ is φ-connected with the field $[Y_1, Y_2]$.* [**Hint.** The fields $X \in \mathfrak{a}\mathscr{X}$ and $Y \in \mathfrak{a}\mathscr{Y}$ are φ-connected if and only if the equality $X(f \circ \varphi) = Yf \circ \varphi$ holds for any function $f \in \mathsf{F}\mathscr{Y}$.]

Now we apply the general construction we have presented to the special case of diffeomorphisms of the form φ_t.

Let S be an arbitrary tensor field of type (a, b) on a manifold $\mathscr{X}$, and let $p \in \mathscr{X}$.

By definition $p \in O_t$ for $|t| < \delta(p)$, and hence for any t with $|t| < \delta(p)$ a tensor $(\varphi_t^*S)_p$ is defined at p, (where S denotes a restriction of a tensor field S on O_t) and hence a tensor $(\varphi_t^*S)_p - S_p$ as well. We put

$$(\pounds_X S)_p = \lim_{t \to 0} \frac{(\varphi_t^*S)_p - S_p}{t}, \tag{12}$$

where X is a vector field generated by a flow $\{\varphi_t\}$.

Since the point p was an arbitrary point of a manifold, tensors (12) make up a tensor field of type (a, b) $\pounds_X S$ on a manifold $\mathscr{X}$. We shall show below by calculating the components of the field in an arbitrary chart (U, h) that the *field $\pounds_X S$ is smooth.*

Definition 4. A field $\pounds_X S$ is called a *Lie derivative* of a tensor field with respect to a vector field X.

It is easy to see that *for every vector field X the mapping $\pounds_X$ is a differentiation of an algebra of tensor fields on a manifold $\mathscr{X}$*, i.e. it is linear, and for any tensor fields S and T we have

$$\pounds_X (S \otimes T) = \pounds_X S \otimes T + S \otimes \pounds_X T. \tag{13}$$

Indeed, the statement about linearity is obvious, and to prove (13) it suffices to notice that by (11)

$$\varphi_t^* (S \otimes T) - S \otimes T = (\varphi_t^* S - S) \otimes \varphi_t^* T \\ + S \otimes (\varphi_t^* T - T)$$

for any t. □

Problem 6. Prove that the *operation* $\pounds_X$ *commutes with the contraction operation of tensor fields* (for any pair of indices).

If the field S is a smooth function f, then by (12)

$$(\pounds_X f)(p) = \lim_{t\to 0} \frac{f(\varphi_t(p)) - f(p)}{t}$$

for any point $p \in \mathscr{X}$. Let $(U, h) = (U, x^1, \ldots, x^n)$ be a chart of a manifold $\mathscr{X}$ such that $p \in U$, and let $f = f(x^1, \ldots, x^n)$ on U. Then if $x^i = x^i(t)$, $i = 1, \ldots, n$, are parametric equations of an integral curve $t \mapsto \varphi_t(p)$ of a field X in a chart (U, h), then

$$\lim_{t\to 0} \frac{f(\varphi_t(p)) - f(p)}{t}$$

$$= \lim_{t\to 0} \frac{f(x^1(t), \ldots, x^n(t)) - f(x^1(0), \ldots, x^n(0))}{t}$$

$$= \frac{df(x^1(t), \ldots, x^n(t))}{dt}\bigg|_{t=0}$$

$$= \left(\frac{\partial f}{\partial x^i}\right)_p \frac{dx^i(t)}{dt}\bigg|_{t=0} = \left(\frac{\partial f}{\partial x^i}\right)_p X_p^i = (Xf)(p),$$

where X^i, $i = 1, \ldots, n$, are the components of a vector field in a chart (U, h).

This proves that $\pounds_X f = Xf$ on U and hence, in view of (U, h) being arbitrary, that

$$\pounds_X f = Xf \quad \text{on} \quad \mathscr{X}. \tag{14}$$

Thus the *operation* $\pounds_X$ *is the generalization of the operation* X *from functions to arbitrary tensor fields.*

By (13) it follows, in particular, that

$$\pounds_X (fS) = Xf \cdot S + f \pounds_X S \tag{15}$$

for any function f and any tensor field S.

Now let the field S be a covector field α and hence by (10)

$$(\varphi_t^*\alpha)_p = (d\varphi)_p^*\alpha_{\varphi(p)}.$$

Consequently, the value $(\varphi_t^*\alpha)_p A$ of a covector $\varphi_t^*\alpha$ on a vector $A \in \mathrm{T}_p\mathscr{X}$ is expressed as follows:

$$(\varphi_t^*\alpha)_p A = \alpha_{\varphi(p)}((d\varphi_t)_p A).$$

If therefore in a chart (U, h)

$$A = a^i\left(\frac{\partial}{\partial x^i}\right)_p \text{ and } \alpha = \alpha_j\, dx^j$$

and the mapping φ_t is given by functions

$$y^k = \varphi_t^k(x^1, \ldots, x^n),$$

then

$$\begin{aligned}(\varphi_t^*\alpha)_p A\\ &= [\alpha_j(\varphi_t(p))(dx^j)_{\varphi_t(p)}]\left(a^i\left(\frac{\partial\varphi_t^k}{\partial x^i}\right)_p\left(\frac{\partial}{\partial x^k}\right)_{\varphi_t(p)}\right)\\ &= \alpha(\varphi_t(p))\, a^i\left(\frac{\partial\varphi_t^j}{\partial x^i}\right)_p\end{aligned}$$

and, in particular,

$$(\varphi_t^*\alpha)\, p\left(\frac{\partial}{\partial x^i}\right)_p = \alpha_j(\varphi_t(p))\left(\frac{\partial\varphi_t^j}{\partial x^i}\right)_p.$$

Since for any covector $\xi \in \mathrm{T}_p^*\mathscr{X}$

$$\xi = \xi\left(\frac{\partial}{\partial x^i}\right)_p (dx^i)_p,$$

this proves that

$$(\varphi_t^*\alpha)_p = (\alpha_j \circ \varphi_t)(p)\left(\frac{\partial\varphi_t^j}{\partial x^i}\right)_p (dx^i)_p.$$

Therefore

$$(\varphi_t^*\alpha)_p - \alpha_p = \left[(\alpha_j \circ \varphi_t)(p)\left(\frac{\partial\varphi_t^j}{\partial x^i}\right)_p - \alpha_i(p)\right](dx^i)_p$$

and hence

$$(\pounds_X\alpha)_p = \beta_i(p)(dx^i)_p,$$

where

$$\beta_i(p)=\lim_{t\to 0}\frac{(\alpha_j\circ\varphi_t)(p)\left(\dfrac{\partial\varphi_t^j}{\partial x^i}\right)_p-\alpha_i(p)}{t}.$$

Let, in particular, $\alpha = dx^k$, i.e. $\alpha_j = \delta_j^k$. Then $(\alpha_j \circ \varphi_t) \times (p) = \delta_j^k$ for any t and hence

$$\beta_i(p)=\lim_{t\to 0}\frac{\left(\dfrac{\partial\varphi_t^k}{\partial x^i}\right)_p-\delta_i^k}{t}=\frac{\partial}{\partial t}\left(\frac{\partial\varphi_t^k}{dx^i}\right)_p\Bigg|_{t=0}$$

$$=\left(\frac{\partial}{\partial x^i}\left(\frac{\partial\varphi_t^k}{\partial t}\bigg|_{t=0}\right)\right)_p.$$

(Notice that $\delta_i^k=\left(\dfrac{\partial\varphi_0^k}{\partial x^i}\right)_p$, since $\varphi_0^k(x^1, \ldots, x^n)=x^k$.) But if, as above, $x^i = x^i(t)$ are parametric equations of a curve $t \mapsto \varphi_t(p)$ in a chart (U, h), then by definition

$$x^i(t) = \varphi_t^i(x_0^1, \ldots, x_0^n), \quad i = 1, \ldots, n,$$

where $x_0^1, \ldots, x_0^n$ are the coordinates of a point p, and hence

$$\frac{\partial\varphi_t^k}{\partial t}\bigg|_{t=0}=\frac{dx^k(t)}{dt}\bigg|_{t=0}=X^k(p), \quad k=1, \ldots, n,$$

where X^k, $k = 1, \ldots, n$, are the components of a vector field X in the chart (U, h). In the case at hand therefore

$$\beta_i(p)=\left(\frac{\partial X^k}{\partial x^i}\right)_p.$$

This proves that

$$\pounds_X\, dx^k=\frac{\partial X^k}{\partial x^i}dx^i \text{ on } U,$$

and hence (see (15))

$$\pounds_X(\alpha_k\, dx^k)=X\alpha_k\, dx^k+\alpha_k\frac{\partial X^k}{\partial x^i}dx^i,$$

i.e.

(16) $$\pounds_X\alpha=\left(X\alpha_i+\alpha_k\frac{\partial X^k}{\partial x^i}\right)dx^i \text{ on } U,$$

which can be written in a more symmetric form:

$$\pounds_X\alpha = \left(X^k \frac{\partial \alpha_i}{\partial x^k} + \alpha_k \frac{\partial X^k}{\partial x^i}\right) dx^i \text{ on } U.$$

In particular, we see in complete agreement with the above general statement that the covector field $\pounds_X\alpha$ is smooth.

A Lie derivative $\pounds_X Y$ of a vector field Y with respect to a vector field X can be obtained by similar computation. It is easier, however, to consider a tensor field $\alpha \otimes Y$ for an arbitrary covector field α on a manifold $\mathscr{X}$ and make use first of the fact that by general formula (11)

$$\pounds_X (\alpha \otimes Y) = \pounds_X\alpha \otimes Y + \alpha \otimes \pounds_X Y$$

and second that the operation $\pounds_X$ commutes with the contraction operation of tensor fields for any pair of indices (see Problem 5 above). Since contraction of a field $\alpha \otimes Y$ obviously results in a smooth function

$$\alpha (Y)\colon p \mapsto \alpha_p (Y_p),$$

it follows that

$$\pounds_X [\alpha (Y)] = (\pounds_X\alpha) (Y) + \alpha (\pounds_X Y). \tag{17}$$

In an arbitrary chart (U, h) the function $\alpha (Y)$ can obviously be expressed as follows:

$$\alpha (Y) = \alpha_i Y^i \quad \text{on} \quad U,$$

where α_i and Y^i, $i = 1, \ldots, n$, are the components of a field α and a field Y, respectively, and hence, by formulas (14) and (16) already proved, formula (17) assumes the form

$$X [\alpha_i Y^i] = \left(X\alpha_i + \alpha_k \frac{\partial X^k}{\partial x^i}\right) Y^i + \alpha_i Z^i \text{ on } U,$$

where Z^i, $i = 1, \ldots, n$, are the components of a vector field $\pounds_X Y$ in a chart (U, h).

In particular, it follows for $\alpha_i = \delta_i^j$ that

$$XY^j = \frac{\partial X^j}{\partial i} Y^i + Z^j,$$

i.e.,

$$Z^j = XY^j - \frac{\partial X^j}{\partial x^i} Y^i = X^i \frac{\partial Y^j}{\partial x^i} - \frac{\partial X^j}{\partial x^i} Y^i.$$

Comparing this result with (22) of the preceding lecture, we immediately see that $Z^j = [X, Y]^j$ and consequently

$$\pounds_X Y = [X, Y]. \tag{18}$$

In an arbitrary Lie algebra $\mathfrak{g}$ the mapping

$$x \mapsto [a, x], \quad a, x \in \mathfrak{g},$$

is designated ad a. Using this notation we can rewrite (18) as

$$\pounds_X = \operatorname{ad} X \quad \text{on} \quad \mathfrak{a}X.$$

We have thus learnt to compute the operation $\pounds_X$ on functions, on covector fields, and on vector fields. Since an arbitrary tensor field S in every chart (U, h) can be expressed as follows:

$$S = S^{j_1 \dots j_b}_{i_1 \dots i_a}\, dx^{i_1} \otimes \dots \otimes dx^{i_a} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{j_b}},$$

we can using (11) (and its special case (15)) to calculate in (U, h) all components $(\pounds_X S)^{j_1 \dots j_b}_{i_1 \dots i_a}$ of a field $\pounds_X S$. Without writing out an explicitly corresponding formula, we can a priori say that it yields an expression for the components $(\pounds_X S)^{j_1 \dots j_b}_{i_1 \dots i_a}$ as a sum of the products of the components of the fields S, X and their derivatives. These components are therefore smooth functions, and hence the field $\pounds_X S$ is smooth.

Problem 7. Prove that

$$\begin{aligned}(\pounds_X S)^{j_1 \dots j_b}_{i_1 \dots i_a} = {} & \frac{\partial S^{j_1 \dots j_b}_{i_1 \dots i_a}}{\partial x^k} X^k + S^{j_1 j_2 \dots j_b}_{k i_2 \dots i_a} \frac{\partial X^k}{\partial x^{i_1}} \\ & + S^{j_1 j_2 \dots j_b}_{i_1 k \dots i_a} \frac{\partial X^k}{\partial x^{i_2}} \\ & + \dots + S^{j_1 \dots j_{b-1} j_b}_{i_1 \dots i_{a-1} k} \frac{\partial X^k}{\partial x^{i_a}} \\ & - S^{k j_2 \dots j_b}_{i_1 i_2 \dots i_a} \frac{\partial X^{j_1}}{\partial x^k} - S^{j_1 k \dots j_b}_{i_1 i_2 \dots i_a} \frac{\partial X^{j_2}}{\partial x^k} \\ & - \dots - S^{j_1 \dots j_{b-1} k}_{i_1 \dots i_{a-1} i_a} \frac{\partial X^{j_b}}{\partial x^k}.\end{aligned}$$

Lecture 18

Linear differential forms · Differential forms of arbitrary degree · Differential forms as functionals of vector fields · The inner product of a vector field and a differential form · Pullback of a differential form via a smooth mapping

We now proceed to study covector fields on $\mathscr{X}$. Every such field α in an arbitrary chart $(U, x^1, \ldots, x^n)$ can be written as

$$\alpha = \alpha_i \, dx^i \quad \text{on} \quad U$$

(see (13) of Lecture 16), i.e. as a linear form of differentials $dx^1, \ldots, dx^n$ of local coordinates. For this reason covector fields are generally called *linear differential forms* (and functions α_i, $i = 1, \ldots, n$, are their *coefficients* in a chart (U, h)). A vector space $\mathrm{T}_1\mathscr{X}$ of linear differential forms is also designated $\Omega^1\mathscr{X}$.

For any linear differential form α and any vector field X the formula

$$\alpha (X) (p) = \alpha_p (X_p), \quad p \in \mathscr{X},$$

defines a function $\alpha (X)$ on $\mathscr{X}$. This function is designated $\langle\alpha, X\rangle$, $i_X\alpha$ and $X \lrcorner\, \alpha$ and called the *inner product* of a form α and a field X. In every chart $(U, h) = (U, x^1, \ldots, x^n)$ the function $\alpha (X)$ is expressed as

$$\alpha (X) = \alpha_i X^i, \tag{1}$$

where X^i are the components of a field X and α_i are the coefficients of a form α in (U, h), from which it follows directly that the *function $\alpha (X)$ is smooth on $\mathscr{X}$*. [In Lecture 17 we have already considered this function *ad hoc* when computing the Lie derivative of a vector field.]

For any differential form $\alpha \in \Omega^1\mathscr{X}$ the formula

$$X \mapsto \alpha(X)$$

defines some mapping

$$\alpha\colon \mathfrak{a}\mathscr{X} \to \mathsf{F}\mathscr{X} \tag{2}$$

which is obviously a morphism of $\mathsf{F}\mathscr{X}$*-modules*, i.e. satisfies the relation

$$\alpha(fX) = f\alpha(X)$$

for any function $f \in \mathsf{F}\mathscr{X}$ and any field $X \in \mathfrak{a}\mathscr{X}$.

We shall say that morphism (2) is generated by a differential form α.

Proposition 1. *For any Hausdorff smooth manifold* $\mathscr{X}$ *the correspondence*

$$\textit{form } \alpha \Rightarrow \textit{morphism } (2) \tag{3}$$

is an isomorphism of a vector space $\Omega^1\mathscr{X}$ *onto a vector space* $\mathrm{Hom}_{\mathsf{F}\mathscr{X}}(\mathfrak{a}\mathscr{X}, \mathsf{F}\mathscr{X})$ *of morphisms* (2).

Before proving this proposition we establish some simple lemmas on morphisms of the form (2). In these lemmas $\mathscr{X}$ will be assumed to be a Hausdorff manifold.

We shall say that vector fields X and Y *coincide near a point* $p \in \mathscr{X}$, if they coincide on some neighbourhood of that point. (Cf. a similar definition for functions in Lecture 16.)

Lemma 1 (the property that morphisms $\mathfrak{a}\mathscr{X} \to \mathsf{F}\mathscr{X}$ are local). *If vector fields* X *and* Y *coincide near a point* $p \in \mathscr{X}$, *then for any morphism* (2) *functions* $\alpha(X)$ *and* $\alpha(Y)$ *also coincide near the point* p.

Proof. (Cf. the proof of Corollary 2 to Lemma 1 of Lecture 16.) Let $X = Y$ on a neighbourhood U of a point p, and let φ be a smooth function $\mathscr{X} \to \mathbb{R}$ such that $\varphi = 0$ outside U and $\varphi = 1$ on some neighbourhood W of p contained in U (see Corollary 1 to Lemma 1 of Lecture 16). Then the field $\varphi\cdot(X - Y)$ is zero on $\mathscr{X}$, i.e. is the zero of a vector space $\mathfrak{a}\mathscr{X}$. Since the morphism α is linear,

$$\alpha[\varphi\cdot(X - Y)] = 0$$

and $\varphi\cdot\alpha(X - Y) = 0$. Since $\varphi = 1$ on W, this proves hat $\alpha(X - Y) = 0$ on W and hence $\alpha(X) = \alpha(Y)$ on W. □

Lemma 2. *Let U be open in $\mathscr{X}$, and let $p_0 \in U$. For any vector field X' on U there is then a vector field X on $\mathscr{X}$ such* hat *$X' = X$ near p_0.*

Proof. (Cf. the proof of Lemma 1 of Lecture 16.) By Proposition 2 of Lecture 14 there are open sets V and W n $\mathscr{X}$ such that

$$p_0 \in W, \quad \overline{W} \subset V, \quad \overline{V} \subset U,$$

and for a pair (V, W) there is a Uryson function φ. For an arbitrary point $p \in \mathscr{X}$ we put

$$X_p = \begin{cases} \varphi(p) X'_p & \text{if } p \in U, \\ 0 & \text{if } p \notin U. \end{cases}$$

It is clear that the field $X\colon p \mapsto X_p$ is smooth on $\mathscr{X}$ and coincides with the field X' on W. □

Corollary 1. *For any point p_0 of a manifold $\mathscr{X}$ and any vector $A \in \mathbf{T}_{p_0}\mathscr{X}$ there is a vector field X on $\mathscr{X}$ such that*

$$X_{p_0} = A.$$

Proof. Let $(U, h) = (U, x^1, \ldots, x^n)$ be a chart such that $p_0 \in U$, and let

$$A = a^i \left(\frac{\partial}{\partial x^i}\right)_{p_0}.$$

Define on U a vector field X' putting

$$X'_p = a^i \left(\frac{\partial}{\partial x^i}\right)_p$$

for every point $p \in U$. By Lemma 2 there is a field X on $\mathscr{X}$ such that $X' = X$ near the point p_0. In particular, $X_{p_0} = X'_{p_0} = A$. □

Corollary 2. *For any open set $U \subset \mathscr{X}$ every morphism α: $\mathsf{a}\mathscr{X} \to \mathsf{F}\mathscr{X}$ induces a unique morphism*

$$\alpha_U\colon \mathsf{a}U \to \mathsf{F}U \tag{4}$$

for which there is a commutative diagram

$$\begin{tikzcd} \mathsf{a}\mathscr{X} \arrow[r] \arrow[d, "\alpha"'] & \mathsf{a}U \arrow[d, "\alpha_U"] \\ \mathsf{F}\mathscr{X} \arrow[r] & \mathsf{F}U \end{tikzcd} \tag{5}$$

whose horizontal arrows are restriction mappings, i.e. such that

$$\alpha(X)|_U = \alpha_U(X|_U)$$

for any vector field X on $\mathscr{X}$. (Cf. Proposition 1 of Lecture 16.)

Proof. Let a morphism α_U exist, and let X be an arbitrary vector field on U. By Lemma 2, for any point $p \in U$ on $\mathscr{X}$ there is a vector field X' coinciding near p with X. If $X' = X$ on a neighbourhood W of p, then

$$\alpha_U(X)|_W = \alpha(X')|_W$$

and, in particular,

$$\alpha_U(X)(p) = \alpha(X')(p). \tag{6}$$

Moreover, by virtue of the property that a morphism α is local for any two fields X' and X'' on $\mathscr{X}$ coinciding near p with X there is an equation

$$\alpha(X')(p) = \alpha(X'')(p) \tag{6'}$$

which shows that the right-hand side of (6) is independent of the choice of the field X. This proves that the morphism α_U is unique.

To prove that the morphism α_U exists we define for any vector field X on U a function $\alpha_U(X)$ by formula (6). According to (6′) this definition is correct. Moreover, if $X' = X$ on a neighbourhood W of p, then $\alpha_U(X) = \alpha(X')$ on W, from which it follows that the function $\alpha_U(X)$ is smooth on U, and hence the formula $\alpha_U\colon X \mapsto \alpha_U(X)$ defines some mapping

$$\alpha_U\colon \mathfrak{a}U \to \mathsf{F}U.$$

If now $X \in \mathfrak{a}U$ and $f \in \mathsf{F}U$ and $X_1 \in \mathfrak{a}\mathscr{X}$ and $f_1 \in \mathsf{F}\mathscr{X}$, with $X = X_1$ and $f = f_1$ near p, then $fX = f_1X_1$ near p and therefore

$$\begin{aligned}\alpha_U(fX)(p) &= \alpha(f_1X_1)(p) = (f_1\alpha(X))(p)\\ &= f_1(p)\cdot\alpha(X_1)(p) = f(p)\cdot\alpha_U(X)(p)\\ &= [f\alpha_U(X)](p).\end{aligned}$$

Consequently, $\alpha_U(fX) = f\alpha_U(X)$ and hence α_U is a morphism of modules.

Finally, since every field X on $\mathscr{X}$ may serve as a field X' for a field $X|_U$ (with respect to an arbitrary point $p \in U$), we have

$$\alpha_U (X|_U)(p) = \alpha(X)(p)$$

at every point $p \in U$. Consequently, diagram (5) is commutative. □

Lemma 3. *In any chart* $(U, h) = (U, x^1, \ldots, x^n)$ *every morphism* (4) *acts by the formula*

$$\alpha_U (X) = \alpha_i X^i, \tag{7}$$

where α_i, $i = 1, \ldots, n$, *are some functions smooth on* U, *and* X^i, $i = 1, \ldots, n$, *are the components of a vector field* X *in the chart* (U, h).

Proof. Since $X = X^i \frac{\partial}{\partial x^i}$ on U, we have

$$\alpha_U (X) = \alpha_U \left(\frac{\partial}{\partial x^i}\right) X^i = \alpha_i X^i,$$

where $\alpha_i = \alpha\left(\frac{\partial}{\partial x^i}\right)$. □

Corollary 1. *Morphism* (4) *is generated on* U *by a differential form*

$$\alpha_U = \alpha_i \, dx^i. \quad \square \tag{8}$$

Now we are in a position to prove Proposition 1.

Proof of Proposition 1. In order not to get entangled in identifications, in this proof we shall denote morphism (2) generated by a form α by $\hat{\alpha}$. It is clear that mapping (3) is linear. Let $\hat{\alpha} = 0$, i.e. $\alpha_p (X_p) = 0$ for any vector field X on $\mathscr{X}$ and any point $p \in \mathscr{X}$. According to Corollary 1 to Lemma 2 there is a field $X \in \mathsf{a}\mathscr{X}$ for any point $p_0 \in \mathscr{X}$ and every vector $A \in \mathbf{T}_{p_0}\mathscr{X}$, such that $X_{p_0} = A$. Consequently, $\alpha_{p_0} (A) = 0$, i.e. $\alpha_{p_0} = 0$ on $\mathbf{T}_{p_0}\mathscr{X}$. Therefore $\alpha = 0$.

This proves that mapping (3) is monomorphic.

Let β be an arbitrary morphism $\mathsf{a}\mathscr{X} \to \mathsf{F}\mathscr{X}$. By Corollary 1 to Lemma 3, for any chart $(U, h) = (U, x^1, \ldots, x^n)$ the morphism β_U: $\mathsf{a}U \to \mathsf{F}U$ is generated by form (8), where $\alpha_i = \beta\left(\frac{\partial}{\partial x^i}\right)$. If $(U', h') = (U', x^{1'}, \ldots,$

$x^{n'}$) is another chart and the morphism $\beta_{U'}$ on U' is generated by a form $\alpha_{U'} = \alpha_{i'}\, dx^{i'}$, then, the morphism β being local on the intersection $U \cap U'$, there is an equation

$$\alpha_i dx^i = \alpha_{i'} dx^{i'}.$$

This shows that putting

$$\alpha_p = (\alpha_U)_p \quad \text{if} \quad p \in U$$

we correctly define on $\mathscr{X}$ a covector field α having the property that $\alpha\,|_U = \alpha_U$ for any coordinate neighbourhood U. This means that α is smooth and generates a given morphism β (i.e. is such that $\hat{\alpha} = \beta$).

Consequently, mapping (3) is epimorphic. □

In what follows we as a rule identify linear differential forms (covector fields) on $\mathscr{X}$ and morphisms (2) they generate.

Remark 1. Note that *mapping* (3) *is itself a morphism of* $\mathsf{F}\mathscr{X}$*-modules*, i.e. in the notation introduced in proving Proposition 1, for any form $\alpha \in \Omega^1\mathscr{X}$ and any function $f \in \mathsf{F}\mathscr{X}$ there is an equation

$$\widehat{f\alpha} = f\hat{\alpha},$$

where the morphism $f\hat{\alpha}$ is defined, as it is customary in algebra, as follows: $(f\hat{\alpha})(X) = f\hat{\alpha}(X)$.

Of particular importance are tensor fields ω associating every point $p \in \mathscr{X}$ with a *skew-symmetric tensor* ω_p, i.e. (see Lecture II.8) a $(r, 0)$-tensor whose components change sign under any transposition of indices. (The number r is called the *degree* of a field ω.) For any fields θ and ω of skew-symmetric tensors the formula

$$(\theta \wedge \omega)_p = \theta_p \wedge \omega_p,$$

where $\theta_p \wedge \omega_p$ is an outer product of tensors θ_p and ω_p (see Lecture II.8), defines a field $\theta \wedge \omega$ of skew-symmetric tensors, the degree of which is equal to the sum of the degrees of θ and ω. Since the components of the outer product of two tensors are algebraically expressible in terms of the components of cofactors, *for smooth fields* θ *and* ω *the field* $\theta \wedge \omega$ *is smooth.*

All algebraic properties of external multiplication of skew-symmetric tensors (for example, associativity and skew commutativity; see Lecture II.8) remain valid for their fields. It is thus possible not to write brackets in the outer product of any number of fields of skew-symmetric tensors, and for any fields θ and ω of skew-symmetric tensors we have

$$\omega \wedge \theta = (-1)^{rs} \theta \wedge \omega,$$

where r is the degree of the field ω and s is the degree of θ.

When $r = 0$ the field ω is a function f and the outer product $\omega \wedge \theta$ is the usual product $f\theta$ of f by θ.

The familiar expressions of skew-symmetric $(r, 0)$-tensors in terms of outer products of covectors of the conjugate basis (see Proposition 8 of Lecture II.8) show that every field of skew-symmetric tensors on an arbitrary coordinate neighbourhood U can be expressed as

$$\begin{aligned} (9) \qquad \omega &= \frac{1}{r!} \omega_{i_1 \ldots i_r} \, dx^{i_1} \wedge \ldots \wedge dx^{i_r} \\ &= \sum_{1 \leqslant i_1 < \ldots < i_n \leqslant n} \ldots \sum \omega_{i_1 \ldots i_n} \, dx^{i_1} \wedge \ldots \wedge dx^{i_r}, \end{aligned}$$

where $\omega_{i_1 \ldots i_r}$ are $a!$-multiplied components of the field ω in a chart $(U, x^1, \ldots, x^n)$ (which are smooth functions on U). For this reason fields of skew-symmetric tensors are also called *differential forms on a manifold* $\mathscr{X}$.

When $r = 1$ we obtain linear differential forms considered above, and when $r = 0$ we obtain smooth functions on $\mathscr{X}$.

Differential forms of degree $r \geqslant 0$ form a linear subspace $\Omega^r \mathscr{X}$ of the space $\mathsf{T}_r \mathscr{X}$ of all $(r, 0)$-tensor fields. Thus $\Omega^0 \mathscr{X} = \mathsf{F} \mathscr{X}$ and $\Omega^1 \mathscr{X} = \mathsf{T}_1 \mathscr{X}$ (whereas when $r > 1$ the inclusion $\Omega^r \mathscr{X} \subset \mathsf{T}_r \mathscr{X}$ is trivially strict). The symbol $\Omega^1 \mathscr{X}$ has already been used by us earlier.

Notice that

$$\Omega^r \mathscr{X} = 0$$

for any $r > n$.

Interpreting $(r, 0)$-tensors as multilinear functionals of vectors we can associate any differential form ω of degree r and any vector fields $X_1, \ldots, X_r$ with a function $\omega(X_1, \ldots, X_r)$ on $\mathscr{X}$ whose value at $p \in \mathscr{X}$ is given as follows:

$$\omega(X_1, \ldots, X_r)(p) = \omega_p((X_1)_p, \ldots, (X_r)_p).$$

If in a chart $(U, x^1, \ldots, x^n)$

$$\omega = \sum_{1 \leqslant i_1 < \ldots < i_r \leqslant n} \cdots \sum \omega_{i_1 \ldots i_r} dx^{i_1} \wedge \ldots \wedge dx^{i_r}$$

and

$$X_1 = X_1^{i_1} \frac{\partial}{\partial x^{i_1}}, \ \ldots, \ X_r = X_r^{i_r} \frac{\partial}{\partial x^{i_r}},$$

then

$$\omega(X_1, \ldots, X_r) = \sum_{1 \leqslant i_1 < \ldots < i_r \leqslant n} \cdots \sum \omega_{i_1 \ldots i_r} \begin{vmatrix} X_1^{i_1} & \ldots & X_1^{i_r} \\ \ldots & \ldots & \ldots \\ X_r^{i_1} & \ldots & X_r^{i_r} \end{vmatrix} \text{ on } U$$

(see formula (5) of Lecture II.8). Consequently, the *function* $\omega(X_1, \ldots, X_r)$ *is smooth.*

The obtained mapping

$$(10) \qquad \omega\colon \underbrace{\mathfrak{a}\mathscr{X} \times \ldots \times \mathfrak{a}\mathscr{X}}_{r \text{ times}} \to \mathsf{F}\mathscr{X}, \ (X_1, \ldots, X_r) \mapsto \omega(X_1, \ldots, X_r)$$

is obviously $\mathsf{F}\mathscr{X}$*-multilinear*, i.e. it is a morphism of $\mathsf{F}\mathscr{X}$-modules in every independent variable. Moreover, *it is skew-symmetric*, i.e. changes sign under any transposition of independent variables.

Just as for $r = 1$ (see Proposition 1 and Remark 1) *if* $\mathscr{X}$ *is a Hausdorff manifold, then for any* $r \geqslant 1$ *the correspondence*

(11) *form of degree* $r \Rightarrow$ *mapping* (10)

gives an isomorphic mapping of an $\mathsf{F}\mathscr{X}$*-module* $\Omega^r\mathscr{X}$ *onto the* $\mathsf{F}\mathscr{X}$*-module of all skew-symmetric* $\mathsf{F}\mathscr{X}$*-multilinear mappings* (10). The proof repeats practically word for word that of Proposition 1 and will be left to the reader (do not fail to carry it out in detail). In what follows we shall as a rule identify differential forms with the corresponding mappings (10).

Remark 2. Similarly, arbitrary $(r, 0)$-tensor fields on $\mathcal{X}$, $r > 0$, can be identified with (not necessarily, skew-symmetric) $\mathsf{F}\mathcal{X}$-multilinear mappings

$$\underbrace{\mathfrak{a}\mathcal{X} \times \ldots \times \mathfrak{a}\mathcal{X}}_{r \text{ times}} \to \mathsf{F}\mathcal{X}$$

and (r, s)-tensor fields, $r \geqslant 0$, $s \geqslant 0$, with $\mathsf{F}\mathcal{X}$-multilinear mappings of the form

$$(12) \quad S\colon \underbrace{\mathfrak{a}\mathcal{X} \times \ldots \times \mathfrak{a}\mathcal{X}}_{r \text{ times}} \times \underbrace{\Omega^1\mathcal{X} \times \ldots \times \Omega^1\mathcal{X}}_{s \text{ times}} \to \mathsf{F}\mathcal{X}.$$

In particular

$$\mathfrak{a}\mathcal{X} = \operatorname{Hom}_{\mathsf{F}\mathcal{X}} (\Omega^1\mathcal{X}, \mathsf{F}\mathcal{X})$$

(corresponding to a field X is an $\mathsf{F}\mathcal{X}$-linear mapping $i_X\colon \alpha \mapsto X \lrcorner \alpha$). Therefore, when $s = 1$ every field (12) can be interpreted as an $\mathsf{F}\mathcal{X}$-multilinear mapping

$$S\colon \underbrace{\mathfrak{a}\mathcal{X} \times \ldots \times \mathfrak{a}\mathcal{X}}_{r \text{ times}} \to \mathfrak{a}\mathcal{X}$$

associating vector fields $X_1, \ldots, X_r$ with a vector field $S(X_1, \ldots, X_r)$ such that

$$S(X_1, \ldots, X_r) \lrcorner \alpha = S(X_1, \ldots, X_r, \alpha)$$

for any differential form $\alpha \in \Omega^1\mathcal{X}$.

Problem 1. Give a similar interpretation to (12) for $s > 1$.

By virtue of identification (11) every vector field X allows an arbitrary form ω of degree $r > 0$ to be associated with a form $i_X\omega = X \lrcorner \omega$ of degree $r - 1$ whose value on vector fields $X_1, \ldots, X_{r-1}$ is given as follows:

$$(X \lrcorner \omega)(X_1, \ldots, X_{r-1}) = \omega(X, X_1, \ldots, X_{r-1}).$$

A form $X \lrcorner \omega$ is called the *inner product* of a field X and a form ω. [For $r = 1$ it has already been considered at the beginning of this lecture.]

When $r = 0$ (i.e. in the case where the form ω is a function) we assume by definition that $X \lrcorner \omega = 0$ for any field X.

Now notice that in the interpretation of differential forms as mappings (10) the inner product $\theta \wedge \omega$ of a form θ of degree r by a form ω of degree s is given as follows:

$$(13)\quad (\theta \wedge \omega)(X_1, \ldots, X_{r+s}) = \sum_{\sigma} \varepsilon_\sigma \theta(X_{\sigma(1)}, \ldots, X_{\sigma(r)})\, \omega(X_{\sigma(r+1)}, \ldots, X_{\sigma(r+s)}),$$

where the summation is taken over all (r, s)*-permutations* σ, i.e. over all substitutions prescribing the order of the first r and the last s subscripts (prove this !).

It follows that *for the inner product* $X \lrcorner (\theta \wedge \omega)$ *of a vector field* X *by a form* $\theta \wedge \omega$ *we have*

$$(14)\quad X \lrcorner (\theta \wedge \omega) = (X \lrcorner \theta) \wedge \omega + (-1)^r \theta \wedge (X \lrcorner \omega),$$

where r *is the degree of the form* θ. Indeed, by (13) for any fields $X_1, \ldots, X_{r+s-1}$

$$X \lrcorner (\theta \wedge \omega)(X_1, \ldots, X_{r+s-1}) = (\theta \wedge \omega) \times (X, X_1, \ldots, X_{r+s-1}) = \sum_{\sigma} \varepsilon_\sigma \theta(Y_{\sigma(1)}, \ldots, Y_{\sigma(r)}) \times \omega(Y_{\sigma(r+1)}, \ldots, Y_{\sigma(r+s)}),$$

$$((X \lrcorner \theta) \wedge \omega)(X_1, \ldots, X_{r+s-1}) = \sum_{\sigma} \varepsilon_\sigma \theta(Y_1, Y_{\sigma(2)}, \ldots, Y_{\sigma(r)}), \times \omega(Y_{\sigma(r+1)}, \ldots, Y_{\sigma(r+s)}),$$

where $Y_1 = X$, $Y_2 = X_1, \ldots, Y_{r+s} = X_{r+s-1}$ and where in the first sum σ ranges over all (r, s)-permutations, and in the second it ranges only over those for which $\sigma(1) = 1$. Since $\sigma(1) = 1$ or $\sigma(r+1) = 1$ for every (r, s)-permutation σ, it follows that

$$(15)\quad (X \lrcorner (\theta \wedge \omega) - (X \lrcorner \theta) \wedge \omega)(X_1, \ldots, X_{r+s-1}) = \sum_{\sigma} \varepsilon_\sigma \theta(Y_{\sigma(1)}, \ldots, Y_{\sigma(r)}) \times \omega(Y_1, Y_{\sigma(r+2)}, \ldots, Y_{\sigma(r+s)}),$$

where the summation is taken over all permutations σ for which $\sigma(r+1) = 1$.

But every (r, s)-permutation σ with $\sigma(r+1) = 1$ defines by the formula

$$\tau(a) = \begin{cases} \sigma(a) - 1 & \text{if } a = 1, \ldots, r, \\ \sigma(a+1) - 1 & \text{if } a = r+1, \\ & \quad \ldots, r+s-1 \end{cases}$$

an $(r, s-1)$-permutation τ for which

$$Y_{\sigma(1)} = X_{\tau(1)}, \ldots, Y_{\sigma(r)} = X_{\tau(r)},$$
$$Y_{\sigma(r+2)} = X_{\tau(r+1)}, \ldots, Y_{\sigma(r+s)} = X_{\tau(r+s-1)}$$

and

$$\varepsilon_\sigma = (-1)^r \varepsilon_\tau.$$

Therefore the right-hand side of (15) is equal to

$$(-1)^r \sum_\tau \varepsilon_\tau \theta (X_{(\tau)1}, \ldots, X_{\tau(r)}) \, \omega (X, X_{\tau(r+1)}, \ldots, X_{\tau(r+s-1)})$$
$$= (-1)^r (\theta \wedge (X \wedge \omega)) \times (X_1, \ldots, X_{r+s-1}),$$

which proves formula (14). □

A linear operator D carrying forms into forms and satisfying for any forms θ and ω the relation

$$D(\theta \wedge \omega) = D\theta \wedge \omega + (-1)^r \theta \wedge D\omega,$$

where r is the degree of the form θ, is called an *antidifferentiation*. In this terminology the proved statement means that *for any vector field X the operator $i_X = X \lrcorner$ of internal multiplication by X is an antidifferentiation.*

Let $f\colon \mathcal{X} \to \mathcal{Y}$ be an arbitrary smooth mapping and ω be an arbitrary differential form of degree $r \geqslant 0$ on a manifold $\mathcal{Y}$. We associate every point $p \in \mathcal{X}$ with a skew-symmetric tensor $(f^*\omega)_p$ of $\mathbf{T}_p\mathcal{X}$ assuming on vectors $A_1, \ldots, A_r \in \mathbf{T}_p\mathcal{X}$ a value

$$(f^*\omega)_p (A_1, \ldots, A_r) = \omega_q ((df)_p A_1, \ldots, (df)_p A_r),$$

where as ever $q = f(p)$ and

$$(df)_p\colon \mathbf{T}_p\mathcal{X} \to \mathbf{T}_p\mathcal{Y}$$

is the differential of f at p. If

$$(U, h) = (U, x^1, \ldots, x^n) \quad \text{and}$$
$$(V, k) = (V, y^1, \ldots, y^m)$$

are charts of manifolds $\mathscr{X}$ and $\mathscr{Y}$ such that $fU \subset V$ and if

$$y^j = f^j(x^1, \ldots, x^n), \quad j = 1, \ldots, m,$$

are functions expressing in charts (U, h) and (V, k) a mapping f, then

$$(df)_p \left(\frac{\partial}{\partial x^i}\right)_p = \left(\frac{\partial f^j}{\partial x^i}\right)_p \left(\frac{\partial}{\partial y^j}\right)_q, \quad i = 1, \ldots, n$$

and hence

$$(f^*\omega)_p \left(\left(\frac{\partial}{\partial x^{i_1}}\right)_p, \ldots, \left(\frac{\partial}{\partial x^{i_r}}\right)_p\right)$$

$$= \left(\frac{\partial f^{j_1}}{\partial x^{i_1}}\right)_p \cdots \left(\frac{\partial f^{j_r}}{\partial x^{i_r}}\right)_p \omega_q \left(\left(\frac{\partial}{\partial y^{j_1}}\right)_q, \ldots, \left(\frac{\partial}{\partial y^{j_r}}\right)_q\right)$$

for any indices $i_1, \ldots, i_r = 1, \ldots, n$. But by definition

$$\omega_q \left(\left(\frac{\partial}{\partial y^{j_1}}\right)_q, \ldots, \left(\frac{\partial}{\partial y^{j_r}}\right)_q\right) = \omega_{j_1 \ldots j_r}(q)$$
$$= (\omega_{j_1 \ldots j_r} \circ f)(p),$$

where $\omega_{j_1 \ldots j_r}$ are the coefficients of the form ω in a chart (V, k). Introducing by analogy functions

$$(f^*\omega)_{i_1 \ldots i_r}\colon p \mapsto (f^*\omega)_p \left(\left(\frac{\partial}{\partial x^{i_1}}\right)_p, \ldots, \left(\frac{\partial}{\partial x^{i_r}}\right)_p\right)$$

we see that for any indices $i_1, \ldots, i_r = 1, \ldots, n$ we have on a neighbourhood U

(16) $$(f^*\omega)_{i_1 \ldots i_r} = \frac{\partial f^{j_1}}{\partial x^{i_1}} \cdots \frac{\partial f^{j_r}}{\partial x^{i_r}} (\omega_{j_1 \ldots j_r} \circ f)$$

which shows in particular that *functions* $(f^*\omega)_{i_1\ldots i_r}$ *are smooth on* U.

Therefore the formula

$$(17)\qquad f^*\omega=\sum_{1\leqslant i_1<\ldots<i_r\leqslant n}\ldots\sum(f^*\omega)_{i_1\ldots i_r}\,dx^{i_1}\wedge\ldots\wedge dx^{i_r}$$

defines on U a differential form $f^*\omega$.

Comparison of the definitions shows that at every point $p\in U$ this form assumes a value $(f^*\omega)_p$. This proves that the *correspondence*

$$p\mapsto(f^*\omega)_p$$

defines a differential form $f^*\omega$ *on* $\mathscr{X}$. In every chart $(U, x^1, \ldots, x^n)$ the form $f^*\omega$ is expressed by formula (17) (i.e. in other words the coefficients of this form in the chart $(U, x^1, \ldots, x^n)$ are functions (16)).

Definition 1. A form $f^*\omega$ is said to be obtained from a form ω by a *pullback* via a smooth mapping f.

It is clear that the mapping

$$f^*\colon \Omega^r\mathscr{Y}\to\Omega^r\mathscr{X},\quad \omega\mapsto f^*\omega,$$

is linear and commutes with an external multiplication:

$$f^*(\theta\wedge\omega)=f^*\theta\wedge f^*\omega$$

for any forms θ and ω on $\mathscr{Y}$.

Moreover, if $f\colon\mathscr{X}\to\mathscr{Y}$ and $g\colon\mathscr{Y}\to\mathscr{Z}$, then

$$(g\circ f)^*=f^*\circ g^*,$$

and if $f=\mathrm{id}\colon\mathscr{X}\to\mathscr{X}$ is an identity mapping, then $f^*\colon\Omega^r\mathscr{X}\to\Omega^r\mathscr{X}$ is also an identity mapping.

For $r=0$, when the form ω is a smooth function $g\colon\mathscr{Y}\to\mathbb{R}$, we have

$$f^*g=g\circ f.$$

When f is a diffeomorphism φ, the construction of $f^*\omega$ is a special case of the general construction φ^*S of Lecture 17. (Cf. Remark 1 of Lecture 17.)

For an arbitrary submanifold $\mathscr{Y}$ of a manifold $\mathscr{X}$ and the corresponding embedding $\iota\colon\mathscr{Y}\to\mathscr{X}$ the mapping

$$\iota^*\colon\Omega^r\mathscr{X}\to\Omega^r\mathscr{Y}$$

is nothing but a *restriction mapping* carrying a form ω on $\mathscr{X}$ into a form $\omega|_{\mathscr{Y}}$ on $\mathscr{Y}$, for which

$$(\omega|_{\mathscr{Y}})_p(A_1, \ldots, A_r) = \omega_p(A_1, \ldots, A_r)$$

at any point $p \in \mathscr{Y}$ and for any vectors $A_1, \ldots, A_r \in \mathbf{T}_p\mathscr{Y}$ (where the space $\mathbf{T}_p\mathscr{Y}$ is naturally regarded as a subspace of a space $\mathbf{T}_p\mathscr{X}$).

Remark 3. The construction of a form $f^*\omega$ can be immediately carried over to arbitrary $(r, 0)$-tensor fields, $r \geqslant 0$. For every such field S on $\mathscr{Y}$ a field f^*S on $\mathscr{X}$ can be given as follows:

$$(f^*S)_p(A_1, \ldots, A_r) = S_q((df)_p A_1, \ldots, (df)_p A_r),$$

where $p \in \mathscr{X}$, $q = f(p)$ and $A_1, \ldots, A_r \in \mathbf{T}_p\mathscr{X}$, and the components $(f^*S)_{i_1 \ldots i_r}$ of f^*S are expressed in terms of the components $S_{i_1 \ldots i_r}$ of S as follows:

$$(f^*S)_{i_1 \ldots i_r} = \frac{\partial f^{j_1}}{\partial x^{i_1}} \cdots \frac{\partial f^{j_r}}{\partial x^{i_r}} (S_{i_1 \ldots i_r} \circ f).$$

Note that this construction cannot be extended to (r, s)-fields, with $s > 0$, (unless f is a diffeomorphism).

Lecture 19

Exterior differential of a differential form · The Lie derivative of a differential form

---

As we already know from Lecture 12, every smooth function f on a manifold $\mathscr{X}$ defines at any point $p \in \mathscr{X}$ a covector $(df)_p$ acting by the formula

$$(df)_p A = Af, \quad A \in \mathbf{T}_p\mathscr{X}$$

and hence defines a linear differential form

$$df\colon p \mapsto (df)_p$$

called a *differential* of f. In every chart $(U, x^1, \ldots, x^n)$ this form can be expressed as

$$df = \frac{\partial f}{\partial x^i} dx^i$$

and is therefore a smooth form. As a morphism of $\mathsf{F}\mathscr{X}$-modules of $\mathfrak{a}\mathscr{X} \to \mathsf{F}\mathscr{X}$ the form df acts by the formula

$$df(X) = Xf, \quad X \in \mathfrak{a}\mathscr{X}. \tag{1}$$

It is clear that the mapping

$$d\colon \Omega^0\mathscr{X} \to \Omega^1\mathscr{X}, \quad f \mapsto df,$$

is linear and has the property that

$$d(fg) = df \cdot g + f \cdot dg$$

for any two functions f and g.

It turns out that the mapping d can be extended in a natural way to differential forms of any degree.

Proposition 1. *For any smooth manifold $\mathscr{X}$ and any $r \geqslant 0$ there is a unique mapping*

$$d\colon \Omega^r\mathscr{X} \to \Omega^{r+1}\mathscr{X}$$

having the following properties:

1° *The mapping d is linear.*

2° *The mapping d is an antidifferentiation, i.e. for any two differential forms θ and ω there is an equation*

$$d(\theta \wedge \omega) = d\theta \wedge \omega + (-1)^r \theta \wedge d\omega,$$

where r is the degree of the form θ.

3° *For any smooth mapping $f\colon \mathscr{X} \to \mathscr{Y}$ and any form ω on $\mathscr{Y}$ there is an equation*

$$df^*\omega = f^*d\omega.$$

4° *For every function $f \in \Omega^0\mathscr{X}$ the form df is its differential* (1).

5° *If $\omega = df$, where $f \in \Omega^0\mathscr{X}$, then*

$$d\omega = 0.$$

Proof. As always in similar situations, we first prove uniqueness.

Let $(U, h) = (U, x^1, \ldots, x^n)$ be an arbitrary chart of the manifold $\mathscr{X}$ and let

$$\omega\,|_U = \sum_{1\leqslant i_1<\ldots<i_r\leqslant n} \cdots \sum \omega_{i_1\ldots i_r} dx^{i_1} \wedge \ldots \wedge dx^{i_r}.$$

Denoting the form

$$d(\omega\,|_U) = d\omega\,|_U$$

by $d\omega_U$ we immediately get by Properties 1°-5°

$$(2)\quad d\omega_U = \sum_{1\leqslant i_1<\ldots<i_r\leqslant n} \cdots \sum d\omega_{i_1\ldots i_r}\, dx^{i_1} \wedge \ldots \wedge dx^{i_r}$$

$$= \sum_{1\leqslant i_1<\ldots<i_r\leqslant n} \cdots \sum \frac{\partial\omega_{i_1\ldots i_r}}{\partial x^i}\, dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_r}.$$

Consequently, the form $d\omega_U = d\omega\,|_U$ and hence the form $d\omega$ (U being an arbitrary coordinate neighbourhood) can be uniquely defined by the form ω. This means that d is a unique mapping.

To prove the existence of d we define on every coordinate neighbourhood U a form $d\omega_U$ with the aid of formula (2). If $(U, x^1, \ldots, x^n)$ and $(U', x^{1'}, \ldots, x^{n'})$ are two charts of the manifold $\mathscr{X}$ and if

$$\omega = \sum_{1 \leqslant i_1 < \ldots < i_r \leqslant n} \cdot \cdot \sum \omega_{i_1 \ldots i_r} dx^{i_1} \wedge \ldots \wedge dx^{i_r} \text{ on } U$$

and

$$\omega = \sum_{1 \leqslant i'_1 < \ldots < i'_r \leqslant n} \ldots \sum \omega_{i'_1 \ldots i'_r} dx^{i'_1} \wedge \ldots \wedge dx^{i'_r} \text{ on } U,$$

then

$$\omega_{i_1 \ldots i_r} = \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial x^{i'_r}}{\partial x^{i_r}} \omega_{i'_1 \ldots i'_r} \text{ on } U \cap U',$$

and hence

$$\frac{\partial \omega_{i_1 \ldots i_r}}{\partial x^i} = \sum_{k=1}^{r} \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial^2 x^{i'_k}}{\partial x^i \, \partial x^{i_k}} \cdots \frac{\partial x^{i'_r}}{\partial x^{i_r}} \omega_{i'_1 \ldots i'_r}$$
$$+ \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial x^{i'_r}}{\partial x^{i_r}} \frac{\partial \omega_{i'_r \ldots i'_r}}{\partial x^i}.$$

On the other hand, for any function f

$$(3) \qquad \frac{\partial^2 f}{\partial x^i \, \partial x^j} dx^i \wedge dx^j$$
$$= \sum_{i<j} \frac{\partial^2 f}{\partial x^i \, \partial x^j} dx^i \wedge dx^j + \sum_{j<i} \frac{\partial^2 f}{\partial x^i \, \partial x^j} dx^i \wedge dx^j$$
$$= \sum_{i<j} \frac{\partial^2 f}{\partial x^i \, \partial x^j} dx^i \wedge dx^j + \sum_{i<j} \frac{\partial^2 f}{\partial x^j \, \partial x^i} dx^j \wedge dx^i$$
$$= \sum_{i<j} \left(\frac{\partial^2 f}{\partial x^i \, \partial x^j} - \frac{\partial^2 f}{\partial x^j \partial x^i} \right) dx^i \wedge dx^j = 0,$$

since $dx^j \wedge dx^i = -dx^j \wedge dx^i$ and the second partial derivatives are independent of the order of differentiation. Therefore, in particular, for any $k = 1, \ldots, r$

$$\frac{\partial^2 x^{i'_k}}{\partial x^i \, \partial x^{i_k}} dx^i \wedge dx^{i_k} = 0$$

and hence

$$\sum_{k=1}^{r} \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial^2 x^{i'_k}}{\partial x^i \, \partial x^{i_k}} \cdots \frac{\partial x^{i'_r}}{\partial x^{i_r}} \omega_{i'_1 \ldots i'_r} dx^i$$
$$\wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k} \wedge \ldots \wedge dx^{i_r} = 0.$$

Consequently, on $U \cap U'$

$$\frac{\partial \omega_{i_1 \ldots i_r}}{\partial x^i} dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_r}$$
$$= \frac{\partial x^{i'_1}}{\partial x^{i_1}} \cdots \frac{\partial x^{i'_r}}{\partial x^{i_1}} \frac{\partial \omega_{i'_1 \ldots i'_r}}{\partial x^r} dx^i$$
$$\wedge dx^{i_1} \wedge \ldots \wedge dx^{i_r}$$
$$= \left(\frac{\partial \omega_{i'_1 \ldots i'_r}}{\partial x^i} dx^i \right) \wedge \left(\frac{\partial x^{i'_1}}{\partial x^{i_1}} dx^{i_1} \right)$$
$$\wedge \ldots \wedge \left(\frac{\partial x^{i'_r}}{\partial x^{i_r}} dx^{i_r} \right)$$
$$= d\omega_{i'_1 \ldots i'_r} \wedge dx^{i'_1} \wedge \ldots \wedge dx^{i'_r}$$
$$= \frac{\partial \omega_{i'_1 \ldots i'_r}}{\partial x^{i'}} dx^{i'} \wedge dx^{i'_1} \wedge \ldots \wedge dx^{i'_r},$$

and hence

$$\sum_{1 \leqslant i_1 < \ldots < i_r \leqslant n} \!\!\cdots\!\! \sum \frac{\partial \omega_{i_1 \ldots i_r}}{\partial x^i} dx^i \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_r}$$
$$= \sum_{1 \leqslant i_1 < \ldots < i_r \leqslant n} \!\!\cdots\!\! \sum \frac{\partial \omega_{i'_1 \ldots i'_r}}{\partial x^{i'}} dx^{i'} \wedge dx^{i'_1} \wedge \ldots \wedge dx^{i'_r}$$

on $U \cap U'$, i.e.

$$d\omega_U = d\omega_{U'} \quad \text{on} \quad U \cap U'.$$

Forms $d\omega_U$ are thus compatible on intersections and make up a differential form $d\omega$ of degree $r + 1$ on a manifold $\mathscr{X}$ with the property that

$$d\omega\,|_U = d\omega_U$$

for any coordinate neighbourhood U.

We have thus constructed the mapping

$$d\colon \Omega^r\mathscr{X} \to \Omega^{r+1}\mathscr{X}.$$

It is clear that this mapping has Properties 1° and 4°. In addition, for any function $f \in \mathsf{F}\mathscr{X}$ in every chart $(U, x^1, \ldots, x^n)$ there is by (3) an equation

$$d(df) = d\left(\frac{\partial f}{\partial x^j}\,dx^j\right) = \frac{\partial^2 f}{\partial x^i\,\partial x^j}\,dx^i \wedge dx^j = 0,$$

which proves Property 5°.

It suffices, of course, to verify Property 2° in an arbitrary chart $(U, x^1, \ldots, x^n)$. Moreover, since the operator d is linear, it suffices to verify this property only for "one-term" forms

$$\theta = f\,dx^\alpha \quad\text{and}\quad \omega = g\,dx^\beta,$$

where it is assumed that

$$dx^\alpha = dx^{i_1} \wedge \ldots \wedge dx^{i_r} \text{ and } dx^\beta = dx^{j_1} \wedge \ldots \wedge dx^{j_s}.$$

But for such forms

$$\begin{aligned} d(\theta \wedge \omega) &= d(fg \wedge dx^\alpha \wedge dx^\beta) \\ &= d(fg) \wedge dx^\alpha \wedge dx^\beta \\ &= (df\cdot g + f\cdot dg) \wedge dx^\alpha \wedge dx^\beta \\ &= df \wedge dx^\alpha \wedge g\,dx^\beta + (-1)^r \\ &\quad \times (f \wedge dx^\alpha) \wedge (dg \wedge dx^\beta) \\ &= d\theta \wedge \omega + (-1)^r\,\theta \wedge d\omega, \end{aligned}$$

which proves Property 2°.

Similarly, it suffices to verify Property 3° only on coordinate neighbourhoods (U, h) and (V, k) which satisfy the relation $fU \subset V$. But in this case if

$$\omega = \sum_{1 \leqslant j_1 < \ldots < j_r \leqslant n} \!\!\ldots \sum \omega_{j_1 \ldots j_r}\,dy^{j_1} \wedge \ldots \wedge dy^{j_r} \text{ on } V,$$

then (see formulas (16) and (17) of Lecture 18)

$$d(f^*\omega)=\sum_{1\leqslant i_i<\dots<i_r\leqslant n}\dots\sum d(f^*\omega)_{i_1\dots i_r}\,dx^{i_1}\wedge\dots\wedge dx^{i_r},$$

where

$$\begin{aligned}d(f^*\omega)_{i_1\dots i_r}&=d\left(\frac{\partial f^{j_1}}{\partial x^{i_1}}\dots\frac{\partial f^{j_r}}{\partial x^{i_r}}(\omega_{j_1\dots j_r}\circ f)\right)\\&=\sum_{s=1}^{r}\frac{\partial f^{j_1}}{\partial x^{i_1}}\dots\frac{\partial^2 f^{j_s}}{\partial x^i\,\partial x^{i_s}}\dots\frac{\partial f^{j_r}}{\partial x^{i_r}}\\&\quad\times(\omega_{j_1\dots j_r}\circ f)\,dx^i\\&\quad+\frac{\partial f^{j_1}}{\partial x^{i_1}}\dots\frac{\partial f^{j_r}}{\partial x^{i_r}}\left(\frac{\partial\omega_{j_1\dots j_r}}{\partial y^j}\circ f\right)\frac{\partial f^j}{\partial x^i}\,dx^i,\end{aligned}$$

and hence (see similar computations made above when constructing the operator d)

$$\begin{aligned}&d(f^*\omega)\\&=\sum_{1\leqslant i_1<\dots<i_r\leqslant n}\dots\sum\frac{\partial f^{j_1}}{\partial x^{i_1}}\dots\frac{\partial f^{j_r}}{\partial x^{i_r}}\left(\frac{\partial\omega_{j_1\dots j_r}}{\partial y^j}\circ f\right)\frac{\partial f^j}{\partial x^i}\,dx^i\\&\qquad\wedge dx^{i_1}\wedge\dots\wedge dx^{i_r}.\end{aligned}$$

On the other hand,

$$d\omega=\sum_{1\leqslant j_1<\dots<j_r\leqslant n}\dots\sum\frac{\partial\omega_{j_1\dots j_r}}{\partial y^j}\,dy^j\wedge dy^{j_1}\wedge\dots\wedge dy^{j_r}$$

and

$$\begin{aligned}&f^*(d\omega)\\&=\sum_{1\leqslant i_1<\dots<i_r\leqslant n}\dots\sum\left(\frac{\partial\omega_{j_1\dots j_r}}{\partial y^j}\circ f\right)f^*dy^j\wedge f^*dy^{j_1}\\&\qquad\wedge\dots\wedge f^*dy^{j_r}\\&=\sum_{1\leqslant i_1<\dots<i_r\leqslant n}\dots\sum\frac{\partial f^{j_1}}{\partial x^{i_1}}\dots\frac{\partial f^{j_r}}{\partial x^{i_r}}\left(\frac{\partial\omega_{j_1\dots j_r}}{\partial y^j}\circ f\right)\frac{\partial f^j}{\partial x^i}\,dx^i\\&\qquad\wedge dx^{i_1}\wedge\dots\wedge dx^{i_r}.\end{aligned}$$

Consequently, $d(f^*\omega)=f^*(d\omega)$. □

Definition 1. A form $d\omega$ is called the *exterior differential* of a form ω.

For the coefficients $(d\omega)_{j_1\dots j_{r+1}}$ of $d\omega$ in a chart $(U, h) = (U, x^1, \dots, x^n)$ to be expressed in terms of the coefficients $\omega_{i_1\dots i_r}$ of ω we first notice that in sum (2) we may restrict ourselves to the summation taken only over distinct indices $i, i_1, \dots, i_r$, since if an index i coincides with one of the indices $i_1, \dots, i_r$ (which are all distinct under the hypothesis), then the corresponding term of sum (2) is zero. If, however, all indices $i, i_1, \dots, i_r$ are distinct, we arrange them in increasing order and denote by $j_1, \dots, j_{r+1}$. Thus, if $i = j_a$, where $a = 1, \dots, r+1$, then

$$i_1 = j_1, \dots, i_{a-1} = j_{a-1}, \quad i_a = j_{a+1}, \dots, i_r = j_{r+1}$$

(recall that under the hypothesis $i_1 < \dots < i_r$), with

$$dx^i \wedge dx^{i_1} \wedge \dots \wedge dx^{i_r} = (-1)^{a-1} dx^{j_1} \wedge \dots \wedge dx^{j_a} \wedge \dots \wedge dx^{j_{r+1}}$$

and

$$\frac{\partial \omega_{i_1\dots i_r}}{\partial x^i} = \frac{\partial \omega_{j_1\dots \hat{j}_a \dots j_{r+1}}}{\partial x^{j_a}},$$

where the sign ^ points out that the corresponding index must be omitted. Since for any sequence $j_1, \dots, j_{r+1}$ of increasing indices the index i may turn out to be any one of them, it follows that

$$\sum_{1\leqslant i_1<\dots<i_r\leqslant n} \cdots \sum \frac{\partial \omega_{i_1\dots i_r}}{\partial x^i} dx^i \wedge dx^{i_1} \wedge \dots \wedge dx^{i_r}$$

$$= \sum_{1\leqslant j_1<\dots<j_{r+1}\leqslant n} \cdots \sum \left(\sum_{a=1}^{r+1} (-1)^{a+1} \frac{\partial \omega_{j_1\dots \hat{j}_a \dots j_{r+1}}}{\partial x^{j_a}} \right) \times dx^{j_1} \wedge \dots \wedge dx^{j_{r+1}},$$

i.e. that

$$(4) \quad (d\omega)_{j_1\dots j_{r+1}} = \sum_{a=1}^{r+1} (-1)^{a+1} \frac{\partial \omega_{j_1\dots \hat{j}_a \dots j_{r+1}}}{\partial x^{j_a}}.$$

For example, for a linear form $\alpha = \alpha_i \, dx^i$

$$(d\alpha)_{ij} = \frac{\partial \alpha_j}{\partial x^i} - \frac{\partial \alpha_i}{\partial x^j}, \quad i < j.$$

Proposition 2. *Considered as an* $\mathsf{F}\mathscr{X}$*-multilinear functional of vector fields, the form* $d\omega$ *is given as*

$$\begin{aligned}(5)\quad (d\omega)&(X_1, \ldots, X_{r+1}) \\ &= \sum_{a=1}^{r+1} (-1)^{a+1} X_a \omega (X_1, \ldots, \hat{X}_a, \ldots, X_{r+1}) \\ &\quad + \sum_{a=1}^{r} \sum_{b=a+1}^{r+1} (-1)^{a+b} \omega ([X_a, X_b], \\ &\qquad X_1, \ldots, \hat{X}_a, \ldots \hat{X}_b, \ldots, X_{r+1}),\end{aligned}$$

where the sign $\hat{}$ *points out that the corresponding field must be omitted.*

Proof. Let $\theta(X_1, \ldots, X_{r+1})$ be the right-hand side of (5). Clearly,

$$\begin{aligned}&\theta(X_1 + X_1', X_2, \ldots, X_{r+1}) \\ &= \theta(X_1, X_2, \ldots, X_{r+1}) + \theta(X_1', X_2, \\ &\quad \ldots, X_{r+1})\end{aligned}$$

for any fields $X_1, X_1', X_2, \ldots, X_{r+1}$. Moreover, for any function $f \in \mathsf{F}\mathscr{X}$

$$\begin{aligned}\theta(fX_1, X_2, \ldots, X_{r+1}) &= fX_1 \omega (X_2, \ldots, X_{r+1}) \\ &+ \sum_{a=2}^{r+1} (-1)^{a+1} X_a \omega (fX_1, \ldots, \hat{X}_a, \ldots, X_{r+1}) \\ &+ \sum_{b=2}^{r+1} (-1)^{b+1} \omega ([fX_1, X_b], X_2, \ldots, \hat{X}_b, \\ &\quad \ldots, X_{r+1}) \\ &+ \sum_{a=2}^{r} \sum_{b=a+1}^{r+1} (-1)^{a+b} \omega ([X_a, X_b], fX_1, \\ &\quad \ldots, \hat{X}_a, \ldots, \hat{X}_b, \ldots, X_{r+1}) \\ &= fX_1 \omega (X_2, \ldots, X_{r+1}) \\ &+ \sum_{a=2}^{r+1} (-1)^{a+1} X_a (f\omega (X_1, \ldots, \hat{X}_a, \ldots, X_{r+1}))\end{aligned}$$

$$+\sum_{b=2}^{r+1}(-1)^{b+1}[f\omega([X_1, X_b], \hat{X}_2, \ldots, \hat{X}_b, \ldots, X_{r+1}) - X_bf\cdot\omega(X_1, \ldots, \hat{X}_b, \ldots, X_{r+1})]$$
$$+\sum_{a=2}^{r}\sum_{b=a+1}^{r+1}(-1)^{a+b}f\omega([X_a, X_b], X_1, \ldots, \hat{X}_a, \ldots, \hat{X}_b, \ldots, X_{r+1})$$
$$= fX_1\omega(X_2, \ldots, X_{r+1})$$
$$+\sum_{a=2}^{r+1}(-1)^{a+1}X_af\cdot\omega(X_1, \ldots, \hat{X}_a, \ldots, X_{r+1})$$
$$+\sum_{a=2}^{r+1}(-1)^{a+1}fX_a\omega(X_1, \ldots, \hat{X}_a, \ldots, X_{r+1})$$
$$+\sum_{b=2}^{r+1}(-1)^{b+1}f\omega([X_1, X_b], X_2, \ldots, \hat{X}_b, \ldots, X_{r+1}) - \sum_{b=2}^{r+1}(-1)^{b+1}X_bf\cdot\omega(X_1, \ldots, \hat{X}_b, \ldots, X_{r+1}) + \sum_{a=2}^{r}\sum_{b=a+1}^{r+1}(-1)^{a+b}f\omega([X_a, X_b], X_1, \ldots, \hat{X}_a, \ldots, \hat{X}_b, \ldots, X_{r+1})$$

(here we make use of the relation $[fX_1, X_b] = f[X_1, X_b] - X_bf\cdot X_1$; see formula (23) of Lecture 16) and hence

$$\theta(fX_1, X_2, \ldots, X_{r+1}) - f\theta(X_1, X_2, \ldots, X_{r+1})$$
$$= \sum_{a=2}^{r+1}(-1)^{a+1}X_af\cdot\omega(X_1, \ldots, \hat{X}_a, \ldots, X_{r+1})$$
$$-\sum_{b=2}^{r+1}(-1)^{b+1}X_bf\cdot\omega(X_1, \ldots, \hat{X}_b, \ldots, X_{r+1})$$
$$= 0.$$

This proves that the *right-hand side of* (5) *is* $\mathsf{F}\mathscr{X}$*-linear in* X_1.

It can similarly be proved (do not fail to carry out the appropriate calculations!) that it is also $\mathsf{F}\mathscr{X}$-linear in the other fields $X_2, \ldots, X_{r+1}$. (This can be proved in another way, which is just a little bit easier, by showing that $\theta(X_1, \ldots, X_{r+1})$ is skew-symmetric in $X_1, \ldots, X_{r+1}$.)

It is therefore sufficient to verify formula (5), in every chart $(U, x^1, \ldots, x^n)$, only for basis vector fields

$$X_1 = \frac{\partial}{\partial x^{j_1}}, \ \ldots, \ X_{r+1} = \frac{\partial}{\partial x^{j_{r+1}}}.$$

But computing both sides of (5) on these fields we obtain the coefficients $(d\omega)_{j_1 \ldots j_{r+1}}$ of the form $d\omega$ on the left and a sum

$$\sum_{a=1}^{r+1} (-1)^{a+1} \frac{\partial}{\partial x^{j_a}} \omega_{j_1 \ldots \hat{j} \ldots j_{r+1}}$$

on the right (the second sum vanishes since $\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0$ for any i and j). By virtue of (4) this proves Proposition 2. ☐

Remark 1. We may try to prove Proposition 2 by verifying that the form $d\omega$ defined by formula (5) has Properties 1°-5° of Proposition 1. Property 1° (linearity of mapping d) and Property 3° (permutability with a transition mapping f^*) are trivial. When $r = 0$ (and $\omega = f$) formula (5) goes into formula (1), which proves Property 4°. When $r = 1$ formula (5) becomes

$$d\omega(X, Y) = X\omega(Y) - Y\omega(X) - \omega([X, Y]). \tag{6}$$

Therefore if $\omega = df$ (and hence $\omega(Z) = Zf$ for any field Z), then

$$d\omega(X, Y) = XYf - YXf - [X, Y]f = 0,$$

and Property 5° is proved. However, a check of the remaining property 2° (carried out using formula (13) of Lecture 18) requires simple but rather cumbersome computations with double sums.

Problem 1. Carry out these computations (and thus reprove Proposition 2).

It turns out that for differential forms the operation of taking a Lie derivative $\mathfrak{L}_X$ can be expressed in terms of inner products on a field X and an operator d.

Proposition 3. *For any vector field X and any differential form ω there is an equation*

$$\pounds_X \omega = X \lrcorner\, d\omega + d\,(X \lrcorner\, \omega). \tag{7}$$

Proof. If ω is a form of degree $r = 0$, i.e. is a function f, then, as we know (formula (14) of Lecture 17), the left-hand side of (7) is equal to Xf. But on the right there remains only the first term $X \lrcorner\, df = df\,(X)$. Since $df\,(X) = Xf$ formula (7) for $r = 0$ is thus proved.

Further, it follows immediately from Property 3° of d that the *operator d is commutative with the operator $\pounds_X$*, i.e.

$$d\pounds_X \omega = \pounds_X\, d\omega \tag{8}$$

for any form ω. In particular,

$$\pounds_X\, df = d\pounds_X f = d\,(Xf)$$

for any function f.

On the other hand, if $\omega = df$, then

$$X \lrcorner\, d\omega + d\,(X \lrcorner\, \omega) = d\,(X \lrcorner\, df) = d\,(Xf).$$

Consequently, formula (7) is true for $\omega = df$ as well.

Since an outer product of forms is obtained by antisymmetrization of their tensor product the analogue of formula (13) of Lecture 17 holds for external multiplication of differential forms as well, i.e.

$$\pounds_X\,(\theta \wedge \omega) = \pounds_X \theta \wedge \omega + \theta \wedge \pounds_X \omega \tag{9}$$

for any forms θ and ω. By Definition 1 of Lecture 16 this means that the *operator $\pounds_X$ is a differentiation of an algebra of forms on $\mathcal{X}$.*

On the contrary, we know that operators d and $X \lrcorner$ are antidifferentiations. But it can easily be seen that for any two antidifferentiations D_1 and D_2 changing degrees by ±1 (or, more generally, by any odd number) the *operator $D_1D_2 + D_2D_1$ is a differentiation.* (Indeed,

$$\begin{aligned}(D_1D_2 + D_2D_1)\,(\theta \wedge \omega) &= D_1\,(D_2\theta \wedge \omega + (-1)^r\,\theta \\ &\quad \wedge D_2\omega) + D_2\,(D_1\theta \wedge \omega + (-1)^r\,\theta \wedge D_1\omega) \\ &= D_1D_2\theta \wedge \omega + (-1)^{r+1}\,D_2\theta \wedge D_1\omega \\ &\quad + (-1)^r\,D_1\theta \wedge D_2\omega + (-1)^{r+r}\,\theta\end{aligned}$$

$$\begin{aligned}&\wedge D_1D_2\omega + D_2D_1\theta \wedge \omega + (-1)^{r\pm 1} D_1\theta\\ &\wedge D_2\omega + (-1)^r D_2\theta \wedge D_1\omega + (-1)^{r+r}\theta\\ &\wedge D_2D_1\omega = (D_1D_2 + D_2D_1)\, \theta \wedge \omega\\ &+ \theta \wedge (D_1D_2 + D_2D_1)\, \omega\end{aligned}$$

for any forms θ and ω.) In particular, a *differentiation is the operator*

$$\omega \mapsto X \lrcorner\, d\omega + d\,(X \lrcorner\, \omega) \tag{10}$$

appearing on the right of relation (7).

Since on any coordinate neighbourhood U an algebra of forms is generated by functions and forms dx^i on which, as we have already seen, operators $\pounds_X$ and (10) coincide and since the fact that two differentiations of some algebra act in the same way on generators obviously implies that these differentiations coincide everywhere, this proves that formula (7) holds on every coordinate neighbourhood U. It is true therefore on the entire manifold $\mathscr{X}$ as well. □

In operator form relation (7) is as follows:

$$\pounds_X = i_X \circ d + d \circ i_X.$$

Problem 2. Prove that

$$[\pounds_X, i_Y] = i_{[X,Y]}$$

for any vector fields X, Y. [**Hint.** Both sides of this formula are antidifferentiations decreasing the degree by -1. Moreover, they are zero on functions and act in the same way on all forms df.]

Problem 3. Prove that

$$[\pounds_X, \pounds_Y] = \pounds_{[X,Y]}$$

on any tensor fields. [**Hint.** Both sides of this formula are differentiations of an algebra of tensor fields acting in the same way on functions, vector fields, and linear differential forms.]

Lecture 20

The de Rham complex and cohomology groups of a smooth manifold · The group $H^0\mathscr{X}$ · Poincaré lemma · The group $H^1\mathbb{S}^2$ · The group $H^1\mathbb{S}^1$ · Computing $H^1\mathbb{S}^1$ using integrals · The group $H^2\mathbb{S}^2$ · Groups $H^1\mathbb{S}^n$ for $n \geqslant 2$ · Groups $H^m\mathbb{S}^n$, $m < n$ · Groups $H^n\mathbb{S}^n$

The equation

$$d\omega = \theta \tag{1}$$

can be written for a given form θ and an unknown form ω in every chart as some system of differential equations with the coefficients of ω. Keeping in mind that differential equations of this sort occur in mathematics and physics we deal in this and subsequent lectures with conditions ensuring the existence of their solutions.

Proposition 1. *For any differential form ω there is an equation*

$$dd\omega = 0.$$

Proof. It suffices to verify this equation in an arbitrary chart and only for forms $\omega = f\,dx^\alpha$, where $dx^\alpha = dx^{i_1} \wedge \ldots \wedge dx^{i_r}$. But by definition

$$d\,(f\,dx^\alpha) = df \wedge dx^\alpha$$

and according to Properties 2° and 5° of the operator d (see Proposition 1 of Lecture 19)

$$dd\,(f\,dx^\alpha) = d\,(df \wedge dx^\alpha) = ddf \wedge dx^\alpha$$
$$- df \wedge ddx^\alpha = 0. \ \square$$

Proposition 1 implies that the equation $d\theta = 0$ is the **necessary condition** for the solvability of equation (1). The determination of conditions under which it is sufficient requires a general study of interrelations between forms $d\omega$ and forms θ for which $d\theta = 0$.

Definition 1. A sequence

$$C^{\bullet}\colon C^0 \xrightarrow{d^0} C^1 \xrightarrow{d^1} \dots \xrightarrow{d^{m-1}} C^m \xrightarrow{d^m} C^{m+1} \to \dots$$

of groups (or, in particular, of vector spaces) and their homomorphisms is said to be a *cochain complex* (the origin of the term will become clear in Lecture 29) if

$$d^m \circ d^{m-1} = 0 \quad \text{for any} \quad m \geqslant 1. \tag{2}$$

It is usual to write d rather than d^m.

The kernel of a homomorphism d^m is designated $Z^mC^{\bullet}$, and its elements are called *cocycles of degree m*. The image $\operatorname{Im} d^{m-1}$ of a homomorphism d^{m-1} is designated $B^mC^{\bullet}$ and its elements are called *coboundaries of degree m*. (When $m = 0$ it is conventionally assumed that $B^0C^{\bullet} = 0$.) Since by condition (2) $B^mC^{\bullet} \subset Z^mC^{\bullet}$, a quotient (or factor) group

$$H^mC^{\bullet} = Z^mC^{\bullet}/B^mC^{\bullet} \tag{3}$$

is defined. It is called the *mth cohomology group* of a complex $C^{\bullet}$ and its elements are *cohomology classes*. Cocycles belonging to the same cohomology class, i.e. such that their difference is a coboundary, are called *cohomologous*.

According to Proposition 1 vector spaces $\Omega^m\mathscr{X}$ and exterior differentials $d\colon \Omega^m\mathscr{X} \to \Omega^{m+1}\mathscr{X}$ make up a cochain complex

$$\Omega^{\bullet}\mathscr{X}\colon \Omega^0\mathscr{X} \xrightarrow{d} \Omega^1\mathscr{X} \to \dots \to \Omega^m\mathscr{X} \xrightarrow{d} \Omega^{m+1}\mathscr{X} \to \dots$$

This complex is called the *de Rham complex* of differential forms of a smooth manifold $\mathscr{X}$. Its cohomology groups are designated $H^m\mathscr{X}$ and called the *de Rham cohomology groups* of a manifold $\mathscr{X}$. (Notice that as a matter of fact they are vector spaces, but the term "vector space of cohomology groups" is not used.)

The coboundaries of a de Rham complex, i.e. forms $d\omega$, are called *exact differentials* or, more briefly, *exact forms*.

The cocycles of a de Rham complex, i.e. forms ω for which $d\omega = 0$, are called *closed forms.* A cohomology class of a closed form ω is designated $[\omega]$.

Thus the necessary condition for the solvability of equation (1) is that the form θ should be closed, and the sufficient condition is that it should be exact. In particular, equation (1') is solvable for any closed form θ of degree m if and only if $H^m\mathscr{X} = 0$. [These statements are of course tautologies. They will become meaningful when we learn to compute the group $H^m\mathscr{X}$ geometrically.]

Let $C^\bullet = \{C^m, d^m\}$ and $D^\bullet = \{D^m, \delta^m\}$ be two cochain complexes.

Definition 2. *By a cochain mapping*

$$\varphi^\bullet\colon C^\bullet \to D^\bullet$$

of a complex $C^\bullet$ into $D^\bullet$ is meant a sequence

$$\varphi^m\colon C^m \to D^m, \quad m \geqslant 0,$$

such that

$$\varphi^{m+1} \circ d^m = \delta^m \circ \varphi^m$$

for any $m \geqslant 0$, i.e. such that the diagram

$$\begin{tikzcd}
C^0 \arrow[r, "d^0"] \arrow[d, "\varphi^0"'] & C^1 \arrow[r] \arrow[d, "\varphi^1"'] & \dots \arrow[r] & C^m \arrow[r, "d^m"] \arrow[d, "\varphi^m"'] & C^{m+1} \arrow[r] \arrow[d, "\varphi^{m+1}"] & \dots \\
D^0 \arrow[r, "\delta^0"] & D^1 \arrow[r] & \dots \arrow[r] & D^m \arrow[r, "\delta^m"] & D^{m+1} \arrow[r] & \dots
\end{tikzcd}$$

is commutative. It is usual to write φ instead of $\varphi^\bullet$ and φ^m.

It is clear that any cochain mapping $\varphi\colon C^\bullet \to D^\bullet$ carries every group $Z^mC^\bullet$ into a group $Z^mD^\bullet$ and every group $B^mC^\bullet$ into a group $B^mD^\bullet$. For any $m \geqslant 0$ therefore it induces some mapping

$$\varphi^*\colon H^mC^\bullet \to H^mD^\bullet.$$

For any cochain mappings $\varphi\colon C^\bullet \to D^\bullet$ and $\psi\colon D^\bullet \to E^\bullet$ there is an equation

$$(\psi \circ \varphi)^* = \psi^* \circ \varphi^*.$$

Moreover, if $\varphi = \mathrm{id}$, then $\varphi^* = \mathrm{id}$. [These properties express the so-called *functoriality* of the correspondence $\varphi \longrightarrow \varphi^*$. We have already encountered them repeatedly.]

Property 3° of the operator d in Proposition 1 of Lecture 19 implies that *for any smooth mapping $f\colon \mathscr{X} \to \mathscr{Y}$ homomorphisms $f^*\colon \Omega^m\mathscr{Y} \to \Omega^m\mathscr{X}$ make up a cochain mapping*

$$f^{\cdot}\colon \Omega^{\cdot}\mathscr{Y} \to \Omega^{\cdot}\mathscr{X}$$

of de Rham complexes. The homomorphisms of cohomology groups induced by this cochain mapping are denoted by the same symbol f^* (this does not lead to any confusion) and called *homomorphisms induced by mapping f.*

Thus by definition

$$f^*[\omega] = [f^*\omega]$$

for any closed form ω of degree m on a manifold $\mathscr{Y}$.

Clearly, if $f = \mathrm{id}\colon \mathscr{X} \to \mathscr{X}$ is an identity mapping, then $f^*\colon H^m\mathscr{X} \to H^m\mathscr{X}$ is also an identity mapping, and for any smooth mappings $f\colon \mathscr{X} \to \mathscr{Y}$ and $g\colon \mathscr{Y} \to \mathscr{Z}$ there is an equation

$$(g \circ f)^* = f^* \circ g^*\colon H^m\mathscr{X} \to H^m\mathscr{Z}.$$

[The correspondence $f \longmapsto f^*$ has the property of being functorial.]

In particular, when $\mathscr{X}$ is a submanifold of a manifold $\mathscr{Y}$ and f is an embedding $\mathscr{X} \to \mathscr{Y}$, the cohomology class $f^*[\omega]$ is designated $[\omega]|_{\mathscr{X}}$ and called a *restriction* of a class $[\omega]$ on $\mathscr{X}$. Thus by definition

$$[\omega]|_{\mathscr{X}} = [\omega|_{\mathscr{X}}]$$

for any cohomology class $[\omega] \in H^m\mathscr{Y}$.

The dimension

$$h^m\mathscr{X} = \dim H^m\mathscr{X}$$

of the group $H^m\mathscr{X}$ (which is, recall, a vector space) is called the mth *Betti number* of a manifold $\mathscr{X}$. (It is also common in the literature to designate it as $b^m\mathscr{X}$.)

Of course, if $\dim \mathscr{X} = n$, then

$$h^m\mathscr{X} = 0 \quad \text{for} \quad m > n$$

(since $\Omega^m\mathscr{X} = 0$ for $m > n$). Of interest are thus only Betti numbers

$$h^0\mathscr{X}, \ldots, h^n\mathscr{X}, \quad n = \dim \mathscr{X}.$$

Notice that the *equation* $h^m\mathscr{X} = 0$ *is necessary and sufficient for equation* (1) *to be solvable on a manifold* $\mathscr{X}$ *for any closed form* θ *of degree* m.

Let us first consider the case $m = 0$. (Although (1) makes sense only for $m > 0$, the case $m = 0$ is still worthy of some attention.)

Let $\pi_0\mathscr{X}$ be the set of all connected components of a smooth manifold $\mathscr{X}$.

Proposition 2. *The vector space* $H^0\mathscr{X}$ *is naturally isomorphic to the vector space* $\mathbb{R}^{\pi_0\mathscr{X}}$ *of all possible mappings* $\pi_0\mathscr{X} \to \mathbb{R}$. *If therefore* $\pi_0\mathscr{X}$ *is a finite set, then the zero Betti number* $h^0\mathscr{X}$ *of a manifold* $\mathscr{X}$ *is equal to the number* $\#\pi_0\mathscr{X}$ *of its connected components*:

$$h^0\mathscr{X} = \#\pi_0\mathscr{X}.$$

In particular, $h^0\mathscr{X} > 0$ *and* $h^0\mathscr{X} = 1$ *if and only if* $\mathscr{X}$ *is a connected manifold.*

We first prove one geometric lemma.

Let $\mathfrak{U} = \{U\}$ be an arbitrary open covering of a manifold $\mathscr{X}$. We shall say that elements U and V of that covering are *chained* if there is a chain of elements U_0, $U_1, \ldots, U_m$ in $\mathfrak{U}$ such that $U_0 = U$, $U_m = V$ and for every $i = 1, \ldots, m$ the sets U_{i-1} and U_i intersect:

$$U_{i-1} \cap U_i \neq \varnothing.$$

Lemma 1. *Every connected component* C *of a manifold* $\mathscr{X}$ *is contained in the union* O *of a subfamily of a covering* $\mathfrak{U}$ *consisting of all the elements of* $\mathfrak{U}$ *chained to a certain element (and hence chained to one another). If the elements of the covering* $\mathfrak{U}$ *are all connected, then* $C = O$.

Proof. Let p_0 be an arbitrary point of the component C and U_0 be an element of $\mathfrak{U}$ such that $p_0 \in U_0$. Further, let O be the union of all elements of $\mathfrak{U}$ chained to U_0. It is clear that O is open (and contains the point p_0). To prove the inclusion $C \subset O$ it therefore suffices to establish that O is simultaneously closed.

Let a point p be in the closure $\bar{O}$ of the set O, and let U be an element of $\mathfrak{U}$ such that $p \in U$. Then $U \cap O \neq \varnothing$ and hence there is an element V of $\mathfrak{U}$ contained in O such that $U \cap V \neq \varnothing$. The element U is therefore chained to U_0, and consequently $U \subset O$. Thus $p \in U \subset O$, and hence O is closed.

If all the elements of $\mathfrak{U}$ are connected, then the set O is connected since so is the union of connected intersecting sets. Therefore $C = O$. □

Proof of Proposition 2. By definition

$$H^0\mathscr{X} = \operatorname{Ker}\{d\colon \Omega^0\mathscr{X} \to \Omega^1\mathscr{X}\},$$

i.e.

$$H^0\mathscr{X} = \{f \in \mathsf{F}\mathscr{X};\ df = 0\}.$$

But the condition $df = 0$ implies that in any chart $(U, h) = (U, x^1, \ldots, x^n)$ there are equations $\frac{\partial f}{\partial x^1} = 0, \ldots, \frac{\partial f}{\partial x^n} = 0$, and consequently if the coordinate neighbourhood U is connected, then $f = \text{const}$ on U. If $f = c_U$ on U and $f = c_V$ on V, then for $U \cap V \neq \varnothing$ $c_U = c_V$.

Bearing this in mind consider an arbitrary covering $\mathfrak{U}$ of a manifold $\mathscr{X}$ consisting of connected coordinate neighbourhoods. According to what has been proved the function f assumes the same value on any two intersecting elements of $\mathfrak{U}$. It therefore assumes the same value on any two chained elements of $\mathfrak{U}$ as well. By Lemma 1 it follows that f is constant on every connected component of $\mathscr{X}$. Since, conversely, every function f constant on a connected component of a manifold $\mathscr{X}$ is smooth on $\mathscr{X}$ and satisfies the relation $df = 0$, this proves that the vector space $H^0\mathscr{X}$ consists of all functions $f\colon \mathscr{X} \to \mathbb{R}$ constant on every connected component of the manifold $\mathscr{X}$ and is hence naturally isomorphic to the vector space of all mappings $\pi_0\mathscr{X} \to \mathbb{R}$. □

Computing numbers $h^m\mathscr{X}$ for $m > 0$ is as a rule quite a difficult problem. In order to set forth the basic principles of solving the problem we must first compare for any manifold $\mathscr{X}$ its cohomology groups with cohomology groups of a manifold $\mathscr{X} \times \mathring{\mathrm{I}}$, where $\mathring{\mathrm{I}} = \mathring{\mathbb{B}}^1$ is an interval $(-1, 1)$.

Recall (see Definition 2 of Lecture 15) that points of the manifold $\mathscr{X} \times \mathring{\mathrm{I}}$ are pairs (p, t), where $p \in \mathscr{X}$, $-1 < t < 1$, and for every chart (U, h) of a manifold $\mathscr{X}$ the pair $(U \times \mathring{\mathrm{I}}, h \times \mathrm{id})$, where

$$(h \times \mathrm{id})(p, t) = (h(p), t) \in \mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R},$$

is a chart of $\mathscr{X} \times \mathring{I}$, charts of this form making up an atlas on $\mathscr{X} \times \mathring{\mathrm{I}}$. If $x^1, \ldots, x^n$ are the local coordinates of a chart (U, h), then the local coordinates of a chart $(U \times \mathring{\mathrm{I}}, h \times \mathrm{id})$ are the functions $x^1 \circ \pi, \ldots, x^n \circ \pi, t$, where π is a projection $(p, t) \mapsto p$. However, one would generally write simply $x^1, \ldots, x^n$ instead of $x^1 \circ \pi, \ldots, x^n \circ \pi$ without fearing any confusion.

If the local coordinates $x^1, \ldots, x^n$ and $x^{1'}, \ldots, x^{n'}$ of charts (U, h) and (U', h') of the manifold $\mathscr{X}$ are connected on $U \cap U'$ by the relations

$$x^{i'} = x^{i'}(x^1, \ldots, x^n), \quad i' = 1, \ldots, n,$$

then the local coordinates $x^1, \ldots, x^n, t$ and $x^{1'}, \ldots, x^{n'}, t$ of charts $(U \times \mathring{\mathrm{I}}, h \times \mathrm{id})$ and $(U' \times \mathring{\mathrm{I}}, h' \times \mathrm{id})$ of a manifold $\mathscr{X} \times \mathring{\mathrm{I}}$ are connected on $(U \cap U') \times \mathring{\mathrm{I}} = (U \times \mathring{\mathrm{I}}) \cap (U' \times \mathring{\mathrm{I}})$ by the relations

$$\begin{aligned} &x^{i'} = x^{i'}(x^1, \ldots, x^n), \quad i' = 1, \ldots, n, \\ &t = t. \end{aligned} \tag{4}$$

It follows that for any chart (U, h) and any point $q_0 = (p_0, t_0) \in U \times \mathring{\mathrm{I}}$ the last vector $\left(\frac{\partial}{\partial t}\right)_{q_0}$ of the basis

$$\left(\frac{\partial}{\partial x^1}\right)_{q_0}, \ \ldots, \ \left(\frac{\partial}{\partial x^n}\right)_{q_0}, \ \left(\frac{\partial}{\partial t}\right)_{q_0}$$

for a space $\mathrm{T}_{q_0}(\mathscr{X} \times \mathring{\mathrm{I}})$ is the *same* for all charts (U, h) with $p_0 \in U$. Therefore the correspondence

$$q_0 \mapsto \left(\frac{\partial}{\partial t}\right)_{q_0}$$

correctly defines on $\mathscr{X} \times \mathring{\mathrm{I}}$ some vector field $\frac{\partial}{\partial t}$.

Similarly, the covector $(dt)_{q_0}$ of the conjugate basis

$$(dx^1)_{q_0}, \ldots, (dx^n)_{q_0}, (dt)_{q_0}$$

for $\mathrm{T}^*_{q_0}(\mathscr{X} \times \mathring{\mathrm{I}})$ is also independent of the choice of the chart (U, h), and hence the correspondence $q_0 \mapsto (dt)_{q_0}$ correctly defines on $\mathscr{X} \times \mathring{\mathrm{I}}$ a linear differential form dt (which is nothing but the differential of a smooth function $(p, t) \mapsto t$).

Among differential forms on $\mathscr{X} \times \mathring{\mathrm{I}}$ there are forms θ in every chart $(U \times \mathring{\mathrm{I}}, x^1, \ldots, x^n, t)$ which are of the form

$$\theta = \sum_{1 \leqslant i_1 < \ldots < i_m \leqslant n} \cdots \sum \theta_{i_1 \ldots i_m} dx^{i_1} \wedge \ldots \wedge dx^{i_m}, \tag{5}$$

where $\theta_{i_1 \ldots i_m}$ are smooth functions of $x^1, \ldots, x^n$ and t. Such forms are said to be *independent of dt*.

For every fixed t, $-1 < t < 1$, form (5) may be assumed to be a differential form on a coordinate neighbourhood U in a manifold $\mathscr{X}$ and it is clear, formulas for the transformation of $x^1, \ldots, x^n$ on $\mathscr{X} \times \mathring{\mathrm{I}}$ and $\mathscr{X}$ being the same, that all such forms (constructed for all possible (U, h)) are compatible on intersections and hence make up some form θ_t on $\mathscr{X}$.

Obviously this establishes a one-to-one correspondence between forms θ on $\mathscr{X} \times \mathring{\mathrm{I}}$ which are independent of dt and families $\{\theta_t\}$ of forms on $\mathscr{X}$ which are *smoothly* dependent on t, $-1 < t < 1$ (i.e. such that in every coordinate neighbourhood their coefficients are smooth functions of t). As a rule, we shall identify θ and $\{\theta_t\}$.

For any form θ independent of dt and for all charts of the form $(U \times \mathring{\mathrm{I}}, x^1, \ldots, x^n, t)$ the forms

$$\int_0^t \theta\, dt = \sum_{1 \leqslant i_1 < \ldots < i_m \leqslant n} \cdots \sum \left(\int_0^t \theta_{i_1 \ldots i_m} dt \right) dx^{i_1} \wedge \ldots \wedge dx^{i_m}$$

are compatible on intersections and hence make up some form $\int_0^t \theta\, dt$ which is also independent of dt.

Similarly defined can be the form $\frac{\partial\theta}{\partial t}$ given in every chart $(U \times \mathring{I}, x^1, \ldots, x^n, t)$ as follows:

$$\frac{\partial\theta}{\partial t} = \sum_{1\leqslant i_1<\ldots<i_m\leqslant n} \cdots \sum \frac{\partial\theta_{i_1\ldots i_m}}{\partial t} dx^{i_1} \wedge \ldots \wedge dx^{i_m},$$

with

$$\frac{\partial}{\partial t}\left(\int_0^t \theta\, dt\right) = \theta$$

for any form θ independent of dt.

Moreover, for any fixed t a form $d\theta_t$ is defined on $\mathscr{X}$. A family of forms $d\theta_t$, $-1 < t < 1$, which is considered as a form on $\mathscr{X} \times \mathring{I}$ is designated $\partial\theta$. The form $\partial\theta$ is of degree $m+1$, where m is the degree of θ, and connected with the exterior differential $d\theta$ of θ as follows:

$$d\theta = \partial\theta + (-1)^m \frac{\partial\theta}{\partial t} \wedge dt. \tag{6}$$

In particular,

$$d\int_0^t \theta\, dt = \partial \int_0^t \theta\, dt + (-1)^m \theta \wedge dt. \tag{7}$$

Among differential forms on $\mathscr{X} \times \mathring{I}$ independent of dt there are in turn forms *independent of* t, i.e. such that in their expressions (5) the coefficients $\theta_{i_1\ldots i_m}$ are independent of t. These forms can be naturally identified with the forms on $\mathscr{X}$ (in fact we have $\pi^*\theta$, where θ is a form on $\mathscr{X}$ and π is a projection $\mathscr{X} \times \mathring{I} \to \mathscr{X}$), so that by virtue of this identification

$$\Omega^m\mathscr{X} \subset \Omega^m(\mathscr{X} \times \mathring{I}) \text{ for all } r \geqslant 0. \tag{8}$$

Forms independent of t are obviously characterized by the relation $\frac{\partial\theta}{\partial t} = 0$, and hence for them $d\theta = \partial\theta$. This means that embeddings (8) make up a cochain mapping

$$\Omega^{\cdot}\mathscr{X} \to \Omega^{\cdot}(\mathscr{X} \times \mathring{\mathrm{I}}) \tag{9}$$

and hence, for any $m \geqslant 0$, induce a homomorphism

$$H^m\mathscr{X} \to H^m\,(\mathscr{X} \times \mathring{\mathrm{I}}) \tag{10}$$

of cohomology groups (which is nothing but a homomorphism π^* induced by π).

Proposition 3. *For any* $m \geqslant 0$ *homomorphism* (10) *is an isomorphism.*

Proof. It is easy to see that an arbitrary form ω on $\mathscr{X} \times \mathring{\mathrm{I}}$ can be uniquely represented as

$$\omega = \theta_1 \wedge dt + \theta_2, \tag{11}$$

where θ_1 and θ_2 are forms independent of dt. (A representation of the form (11) is of course possible in every coordinate neighbourhood U and it is clear that the resulting forms θ_1 and θ_2 are compatible on intersections.) Putting

$$\theta_3 = (-1)^m \int_0^t \theta_1\, dt$$

we get by (7)

$$\omega = d\theta_3 + \theta,$$

where

$$\theta = \theta_2 - \partial\theta_3$$

is a form independent of dt.

This proves, in the case where the form is closed, that

$$[\omega] = [\theta],$$

i.e. any cohomology class of a manifold $\mathscr{X} \times \mathring{\mathrm{I}}$ contains a form θ independent of dt. The form θ is closed together with the form ω, i.e.

$$\partial\theta + \frac{\partial\theta}{\partial t} \wedge dt = 0$$

and, in particular, $\frac{\partial\theta}{\partial t} = 0$. Consequently, the form θ is in fact independent of t as well, i.e. is an image under the mapping π^* of some form on $\mathscr{X}$. This proves that mapping (10) is an epimorphism.

On the other hand, if θ, being independent of t, has the form $d\omega$, where ω is some form on $\mathscr{X} \times \overset{\circ}{\mathrm{I}}$, then representing ω as (11) and noticing that

$$d(\theta_1 \wedge dt + \theta_2) = \left(\partial\theta_1 + (-1)^m \frac{\partial\theta_2}{\partial t}\right) \wedge dt + \partial\theta_2$$

we immediately get

$$\partial\theta_1 + (-1)^m \frac{\partial\theta_2}{\partial t} = 0 \text{ and } \partial\theta_2 = \theta.$$

The equation $\partial\theta_2 = \theta$ holds identically in t and, in particular, at $t = 0$. Since θ is independent of t, this proves that $\partial(\theta_2)_0 = \theta$, where $(\theta_2)_0$ is the value of the form θ_2 at $t = 0$, and hence that $\partial(\theta_2)_0 = \theta$ in $\mathscr{X}$. Consequently, mapping (10) is an isomorphism. □

Let $\overset{\circ}{\mathrm{I}}{}^n$ be an open cube of the space $\mathbb{R}^n$ consisting of points $\mathbf{t} = (t, \ldots, t_n)$ for which $-1 < t_1 < 1, \ldots, -1 < t_n < 1$.

Corollary 1 (Poincaré lemma). *Any closed differential form ω of degree $m > 0$ on a cube $\overset{\circ}{\mathrm{I}}{}^n$ is an exact form, i.e.*

$$h^m \overset{\circ}{\mathrm{I}}{}^n = 0 \textit{ for } m > 0.$$

Proof. Noticing that

$$\overset{\circ}{\mathrm{I}}{}^n = \{\mathrm{pt}\} \times \underbrace{\overset{\circ}{\mathrm{I}} \times \ldots \times \overset{\circ}{\mathrm{I}}}_{n \text{ times}},$$

where $\{\mathrm{pt}\}$ is a zero-dimensional manifold consisting of one point, it suffices to apply Proposition 3 n times. (Clearly, $H^m\{\mathrm{pt}\} = 0$ for $m > 0$.) □

Problem 1. Prove that a *cube* $\overset{\circ}{\mathrm{I}}{}^n$ *is diffeomorphic to an open ball* $\overset{\circ}{\mathbb{B}}{}^n = \{\mathbf{x} \in \mathbb{R}^n;\ |\mathbf{x}| < 1\}$. [Hint. Consider a restriction to $\overset{\circ}{\mathbb{B}}{}^n$ of the homeomorphism $\mathbb{B}^n \to \mathrm{I}^n$ constructed in the proof of Corollary 1 to Theorem 2 of Lecture 9.]

Problem 2. Prove that a *cube* $\overset{\circ}{\mathrm{I}}{}^n$ *is diffeomorphic to* $\mathbb{R}^n$. [Hint. Construct a diffeomorphism $\mathbb{R}^n \to \overset{\circ}{\mathbb{B}}{}^n$.]

It follows that in Corollary 1 a cube $\overset{\circ}{\mathrm{I}}{}^n$ can be replaced either by a ball $\overset{\circ}{\mathbb{B}}{}^n$ or by a space $\mathbb{R}^n$.

Using the Poincaré lemma we now compute cohomology groups $H^m\mathbb{S}^n$ of a sphere $\mathbb{S}^n$ of dimension n.

First let $n = 2$ and $m = 1$.

We know (see Example 9 of Lecture 6 for $n = 2$) that the sphere $\mathbb{S}^2$ has an atlas consisting of six charts which we now denote by (U_k, h_k), where $k = \pm 1, \pm 2, \pm 3$. The supports

$$U_{-3},\ U_{-2},\ U_{-1},\ U_1,\ U_2,\ U_3$$

of these charts are hemispheres consisting of points $(x, y, z) \in \mathbb{S}^2$ for which

$$x < 0,\ y < 0,\ z < 0,\ z > 0,\ y > 0,\ x > 0,$$

respectively. Since each of these supports is diffeomorphic to an open disk $\mathring{\mathbb{B}}^2$; for any closed linear differential form α on $\mathbb{S}^2$ there are by the Poincaré lemma smooth functions

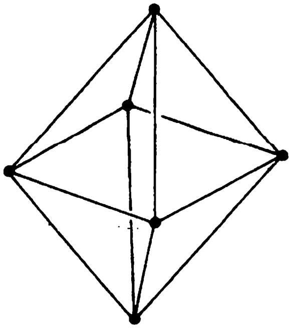
$$f_k\colon U_k \to \mathbb{R},\ k = \pm 1,\ \pm 2,\ \pm 3$$

such that $\alpha = df_k$ on U_k. For any k and l (for which the intersection $U_k \cap U_l$ is not empty) the forms df_k and df_l coincide on $U_k \cap U_l$, and hence, the set $U_k \cap U_l$ being connected, there is a constant c_{kl} such that

$$f_k - f_l = c_{kl} \text{ on } U_k \cap U_l. \tag{12}$$

Conventionally representing neighbourhoods U_k by dots and joining dots corresponding to intersecting neighbourhoods we graphically present a combinatorial scheme of intersections of neighbourhoods of the form

(13)



,

which is nothing but a system of edges of an octahedron, triples of neighbourhoods U_k with a nonempty intersection corresponding to the octahedron faces.

We now prove the following combinatorial lemma.

Lemma 2. *Let every oriented edge kl of octahedron* (13) *be associated with a number* c_{kl}, *the following conditions holding*:

(a) *edges kl and lk are associated with opposite numbers*:

$$c_{lk} = -c_{kl}; \tag{14}$$

(b) *for every face klm there is a relation*

$$c_{kl} + c_{lm} + c_{mk} = 0. \tag{15}$$

Then there are numbers b_k *such that*

$$b_l - b_k = c_{kl} \tag{16}$$

for any edge kl.

Proof. It is necessary to prove that a system of 12 linear equations (16) (in 6 unknowns b_k) is compatible. To do this we first use relations (15) to reduce the number of these equations. For any face *klm* of octahedron (13) relation (15) states that of the three equations (16) corresponding to the edges of that face one is a consequence of the other two. Therefore each of these relations allows the number of equations (16) to be reduced by one. Consequently, using all these relations (and excluding, say, equations corresponding to the edges converging at vertices 1 and —1) we obtain four equations

$$\begin{gathered} b_{-2} - b_{-3} = c_{\bar{3}\bar{2}}, \quad b_3 - b_{-2} = c_{\bar{2}3}, \\ b_2 - b_3 = c_{32}, \quad b_{-3} - b_2 = c_{2\bar{3}} \end{gathered} \tag{17}$$

for the four remaining unknowns b_{-3}, b_{-2}, b_2, b_3 (here and in what follows we write $\bar{1}$, $\bar{2}$, $\bar{3}$ instead of the indices —1, —2, and —3). Putting in these equations, for example, $b_{-3} = 0$ we obtain for the unknowns b_{-2}, b_2, and b_3 the equations

$$b_{-2} = c_{\bar{3}\bar{2}}, \quad b_3 - b_{-2} = c_{\bar{2}3}, \quad b_2 - b_3 = c_{32}$$

from which they can immediately be determined. [The thing is that equations (17) are linearly dependent: their sum is identically zero (check it!). The geometrical reason why in summing these equations all unknowns are cancelled is that equations (17) correspond to edges of octahedron (13) that form a closed path, and the reason why free terms are cancelled is that that path is an edge of a pyramid consisting of faces with vertex at the point 1.]

Thus if conditions (15) (and (14)) hold, equations (16) are in fact compatible.☐

Remark 1. Lemma 2 obviously remains valid, together with the proof, if c_{kl} are elements of an arbitrary (additively written) group (for example, a vector space). Of course, we mean here that b_k are sought in the same group (vector space).

Since numbers c_{kl} given by formula (12) obviously satisfy relations (14) and (15), there are by Lemma 2 numbers b_k, $k = \pm 1, \pm 2, \pm 3$ such that for any edge kl of octahedron (13)

$$f_k - f_l = b_l - b_k \text{ on } U_k \cap U_l,$$

i.e.

$$f_k + b_k = f_l + b_l \text{ on } U_k \cap U_l.$$

It follows that putting

$$f = f_k + b_k \text{ on } U_k$$

we uniquely define on a sphere $\mathbb{S}^2$ a smooth function f, with $df = d(f_k + b_k) = df_k = \alpha$ on U_k and, consequently, $df = \alpha$ on the entire sphere $\mathbb{S}^2$.

This proves that *any closed differential form* α *of degree* 1 *on* $\mathbb{S}^2$ *is an exact form.* This means that $H^1\mathbb{S}^2 = 0$, i.e.

$$h^1\mathbb{S}^2 = 0. \tag{18}$$

Note that besides the Poincaré lemma we have also used the combinatorial lemma 2.

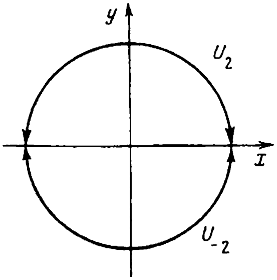
Let us now try to use the same method to compute the group $H^1\mathbb{S}^1$, where $\mathbb{S}^1$ is a circle $x^2 + y^2 = 1$.

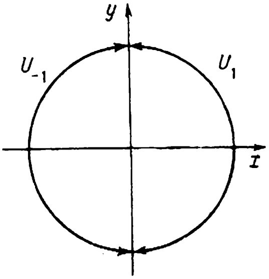
Let

$$U_{-2},\ U_{-1},\ U_1,\ U_2$$

be semicircles characterized by the inequalities

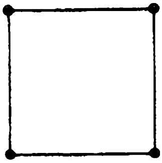
$$x < 0,\ y < 0,\ y > 0,\ x > 0,$$





respectively. The scheme of intersections of semicircles is a square:

(19)



By virtue of the Poincaré lemma, for every linear form ω on S^1 (notice that the form is trivially closed) and any $k = \pm 1, \pm 2$ there are functions f_k on U_k such that

$$\omega = df_k \text{ on } U_k.$$

For any side kl of square (19)

$$d(f_k - f_l) = 0 \text{ on } U_k \cap U_l,$$

and therefore, since the set $U_k \cap U_l$ is connected,

$$f_k - f_l = c_{kl},$$

where c_{kl} are some constants. These as before satisfy relations (14) (but relations (15) are empty for them).

We shall call functions $c: kl \mapsto c_{kl}$ defined on the sides of square (19) and satisfying relation (14) (i.e. such that $c_{lk} = -c_{kl}$) *one-dimensional cocycles of square* (19). (Similarly, functions $c: kl \mapsto c_{kl}$ defined on the edges of octahedron (13) and satisfying relations (14) and (15) are called *one-dimensional cocycles of octahedron* (13), but in the foregoing we went along without this term.) Clearly, cocycles $c\colon kl \mapsto c_{kl}$ form a vector space Z^1 of dimension 4.

Cocycles c for which there is a mapping $b\colon k \mapsto b$ such that

$$b_k - b_l = c_{kl} \tag{20}$$

for any side kl of square (19) are called *coboundaries*. They form a subspace B^1 of Z^1. The corresponding quotient (or factor) space Z^1/B^1 is designated H^1. Its elements (cosets of Z^1 modulo B^1) are called *cohomology classes* of square (19). Cocycles belonging to the same cohomology class are said to be *cohomologous*.

[The terminology suggests that we do in fact deal here with some cochain complex in the sense of Definition 1. We develop this idea in the next lecture.]

The construction of a cocycle $c \in Z^1$ from a form $\omega \in \Omega^1 S^1$ is to a certain extent arbitrary due to the choice of function f_k. But if $\omega = df_k$ and $\omega = d\hat{f}_k$ on U_k, then $d(\hat{f}_k - f_k) = 0$ on U_k and hence $\hat{f}_k = f_k + b_k$, where b_k are some constants. If $c_{kl} = f_k - f_l$ and $\hat{c}_{kl} = \hat{f}_k - \hat{f}_l$ on $U_k \cap U_l$, then

$$\hat{c}_{kl} - c_{kl} = b_k - b_l,$$

i.e. cocycles $c\colon kl \mapsto c_{kl}$ and $\hat{c}\colon kl \mapsto \hat{c}_{kl}$ are cohomologous. This proves that the formula

$$\omega \mapsto [c],$$

where $[c]$ is the cohomology class of a cocycle c, correctly defines some mapping

$$\Omega^1 S^1 \to H^1. \tag{21}$$

If $\omega = df$, then we may take as functions f_k restrictions $f|_{U_k}$ of a function f. Since with such a choice of functions numbers c_{kl} are obviously equal to zero, we see that *on exact forms mapping* (21) *is equal to zero*. It induces therefore some mapping

$$H^1 S^1 \to H^1. \tag{22}$$

It turns out that *mapping* (22) *is a monomorphism*, i.e. if a cocycle $c\colon kl \mapsto c_{kl}$ corresponding to a form $\omega \in \Omega^1\mathbb{S}^1$ is a coboundary, then the form ω is exact. Indeed, if $c_{kl} = f_k - f_l$ on $U_k \cap U_l$ and if $c_{kl} = b_k - b_l$, then $f_k + b_k = f_l + b_l$ on $U_k \cap U_l$, and hence the formula

$$f = f_k + b_k \text{ on } U_k$$

correctly defines on $\mathbb{S}^1$ some function f. Since

$$\omega = df_k = d(f_k + b_k) = df \text{ on } U,$$

we have $\omega = df$ on $\mathbb{S}^1$, and hence the form ω is exact. □

So far we have in fact followed word for word the computation of the group $H^1\mathbb{S}^2$. Continuing the analogy we must now prove the analogue of Lemma 2 (which in the present terms obviously states that for octahedron (13) we have $H^1 = 0$). To this end we must more closely analyze equations (20) (coinciding in form with equations (16) of Lemma 2).

These equations form a system of four

$$\begin{aligned} b_1 - b_2 &= c_{12}, \\ b_2 - b_{-1} &= c_{2\bar{1}}, \\ b_{-1} - b_{-2} &= c_{\bar{1}\bar{2}}, \\ b_{-2} - b_1 &= c_{\bar{2}1} \end{aligned} \tag{23}$$

involving four unknowns b_{-2}, b_{-1}, b_1, b_2 (as before we replace indices of the form $-k$ by $\bar{k}$) and the statement that a cocycle c is a coboundary implies that these equations are compatible.

But adding equations (23) together and putting

$$\operatorname{Ind} c = c_{12} + c_{2\bar{1}} + c_{\bar{1}\bar{2}} + c_{\bar{2}1}$$

we immediately see that for these equations to be compatible it is necessary that $\operatorname{Ind} c = 0$.

The function Ind is a linear mapping $Z^1 \to \mathbb{R}$ and the statement that $\operatorname{Ind} c = 0$ is necessary for equations (23) to be compatible implies that the mapping is zero on a subspace B^1. Therefore the formula

$$\operatorname{Ind}\,[c] = \operatorname{Ind} c,$$

where $[c]$ is the cohomology class of a cocycle c, correctly defines some mapping

$$\text{Ind}: H^1 \to \mathbb{R}. \tag{24}$$

From the first three equations (23) we find successively

$$\begin{aligned} b_2 &= b_1 - c_{12}, \\ b_{-1} &= b_1 - c_{12} - c_{2\bar{1}}, \\ b_{-2} &= b_1 - c_{12} - c_{2\bar{1}} - c_{\bar{1}\bar{2}}. \end{aligned} \tag{25}$$

Thus for any b_1 formulas (25) give a solution of the first three equations (23). If, however, Ind $c = 0$, then the number b_{-2} given by the last formula (25) obviously satisfies the fourth equation (23) as well. Consequently, the condition Ind $c = 0$ is not only necessary but also sufficient for equations (23) to be compatible. For mapping (24) this implies that it is a monomorphism. But since this mapping is obviously epimorphic, it follows that *mapping* (24) *is an isomorphism*, and hence

$$\dim H^1 = 1.$$

[Thus if for an octahedron $H^1 = 0$, then for a square $H^1 \simeq \mathbb{R}$.]

In contrast to the case of the group $H^1\mathbb{S}^2$ the result obtained does not yet allow the group $H^1\mathbb{S}^1$ to be completely defined; it only suggests that either $H^1\mathbb{S}^1 = 0$ or $H^1\mathbb{S}^1 \simeq \mathbb{R}$. It turns out that in fact the latter case holds, i.e.

$$h^1\mathbb{S}^1 = 1. \tag{26}$$

For this to be established, it suffices to produce a form $\omega_0 \in \Omega^1\mathbb{S}^1$ for which the cocycle $c: kl \mapsto c_{kl}$ is not a coboundary, i.e. has the property that Ind $c \neq 0$.

Introduction of coordinates for a circle $\mathbb{S}^1$, known from school, consists in that a number $t \in \mathbb{R}$ is associated with a point $p = (\cos t, \sin t)$ of that circle (geometrically the coordinate t is an angle made by the radius vector of p with the positive direction of the axis of abscissas; for this reason we shall call it an *angular coordinate* on the circle). The coordinate t, appropriately restricted is a local coordinate in the sense of Definition 1 of Lecture 6 on every coordinate neighbourhood U_k (but not on the entire circle S^1, of course). Namely, the coordinate t restricted to

$$(0, \pi), \quad (-\pi, 0), \quad \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \quad \left(\frac{\pi}{2}, \frac{3\pi}{2}\right)$$

is a local coordinate on

$$U_1, U_{-1}, U_2, U_{-2},$$

respectively. This local coordinate is denoted by t_k. Thus

$$(27) \quad \begin{aligned} t_1 &= t_{-2} \text{ on } U_1 \cap U_{-2} & t_{-1} &= t_2 \text{ on } U_{-1} \cap U_2 \\ t_{-2} &= t_{-1} + 2\pi \text{ on } U_{-2} \cap U_{-1} & t_2 &= t_1 \text{ on } U_2 \cap U_1. \end{aligned}$$

Considering the coordinate t on the entire circle S^1 means in fact that by the mapping

$$\mathbb{R} \to S^1, \; t \mapsto (\cos t, \sin t),$$

we go over to the straight line $\mathbb{R}$. Functions on S^1 thus turn out to be periodic (with period 2π) functions on $\mathbb{R}$, and linear differential forms on S^1 to be forms $f(t)\,dt$ on $\mathbb{R}$ with periodic coefficients $f(t)$.

Thus, in particular, a form $\omega_0 = dt$ is defined on S^1 (although t is not a function on S^1), with

$$dt = dt_k \text{ on } U_k.$$

According to relations (27) the cocycle $c = \{c_{kl}\}$ for ω_0 will be expressed as follows:

$$c_{kl} = \begin{cases} 2\pi \text{ if } k = \bar{2},\ l = \bar{1}, \\ -2\pi \text{ if } k = \bar{1};\ l = \bar{2}, \\ 0 \text{ for all the other } k \text{ and } l. \end{cases}$$

Hence $\operatorname{Ind} c = 2\pi \neq 0$.

The form ω_0 can be constructed without resorting to the angular coordinate t.

It is clear that x and y are smooth functions on S^1, and hence differential forms dx and dy are defined on S^1.

Problem 3. Show that

$$\omega_0 = x\,dy - y\,dx$$

and

$$\omega_0 = \begin{cases} -\dfrac{dx}{y} & \text{on } U_1 \text{ and } U_{-1}, \\ \dfrac{dy}{x} & \text{on } U_2 \text{ and } U_{-2}. \end{cases}$$

Rederive from this that Ind $c = 2\pi$.

The group $H^1\mathbb{S}^1$ can also be computed by an easier method which uses no covering $\{U_k\}$ but appeals instead to analytic considerations.

This method is based on associating every form $\omega = f(t)\,dt$ on $\mathbb{S}^1$ with an integral

$$I\omega = \int_0^{2\pi} f(t)\,dt.$$

The resulting mapping

$$\Omega^1\mathbb{S}^1 \to \mathbb{R}, \quad \omega \mapsto I\omega, \tag{28}$$

is obviously linear and epimorphic (since $I\omega_0 \neq 0$ for the form $\omega = dt$). If $\omega = dg$, where g is a function on $\mathbb{S}^1$ (a periodic function on $\mathbb{R}$), i.e. if $f(t) = g'(t)$, then

$$I\omega = g(t)\,\Big|_0^{2\pi} = 0.$$

Conversely, if $I\omega = 0$, then the function

$$g(t) = \int_0^t f(t)\,dt$$

is periodic and is therefore a function on $\mathbb{S}^1$ for which $dg = \omega$.

Thus the kernel of mapping (28) is a subspace of exact forms and hence this mapping induces an isomorphism $H^1\mathbb{S}^1 \to \mathbb{R}$.

Therefore $h^1\mathbb{S}^1 = 1$.

The simplicity and conciseness of this proof demonstrates the strength of methods of integral calculus.

Now let us compute the group $H^2\mathbb{S}^2$.

Again let $\{U_k,\ k = \pm 1, \pm 2, \pm 3\}$ be the cover of $\mathbb{S}^2$ the scheme of mutual intersections of whose elements is octahedron (13), and let ω be an arbitrary (trivially closed) form of degree 2 on the sphere S^2.

Since every set U_k is diffeomorphic to a disk $\mathring{\mathbb{B}}^2$, according to the Poincaré lemma there is for every $k = \pm 1$, ± 2, ± 3 on U_k a form α_k of degree 1 such that $\omega = d\alpha_k$ on U_k. For any k and l (for which $U_k \cap U_l \neq \varnothing$) the form $\alpha_k - \alpha_l$ is closed on $U_k \cap U_l$, and since the set $U_k \cap U_l$ is also diffeomorphic to the disk $\mathring{\mathbb{B}}^2$, again, according to the Poincaré lemma, there is a smooth function $f_{kl}\colon U_k \cap U_l \to \mathbb{R}$ such that

$$\alpha_k - \alpha_l = df_{kl} \text{ on } U_k \cap U_l \tag{29}$$

(we may of course assume without loss of generality that $f_{kl} = -f_{lk}$ for any k and l).

For every three indices k, l, m (for which $U_k \cap U_l \cap U_m \neq \varnothing$) we have

$$d(f_{kl} + f_{lm} + f_{mk}) = \alpha_k - \alpha_l + \alpha_l - \alpha_m + \alpha_m - \\ -\alpha_k = 0,$$

and hence (since $U_k \cap U_l \cap U_m$ is a connected set) there are constants c_{klm} such that

$$f_{kl} + f_{lm} + f_{mk} = c_{klm} \text{ on } U_k \cap U_l \cap U_m$$

(of course, it may be assumed without loss of generality that numbers c_{klm} are skew-symmetrically dependent on indices k, l, and m, i.e. they change sign whenever transposed).

There is of course an element of arbitrariness in this construction. Indeed, first, instead of forms α_k we may take any forms $\hat{\alpha}_k$ having the property that $d\hat{\alpha}_k = d\alpha_k$, i.e. such (we again use the Poincaré lemma here) that $\hat{\alpha}_k = \alpha_k + dg_k$, where $g_k\colon U_k \to \mathbb{R}$ are some smooth functions. Second, functions f_{kl} can be similarly replaced by functions $f_{kl} + b_{kl}$, where b_{kl} are arbitrary constants (satisfying the relation $b_{lk} = -b_{kl}$). Allowing for both possibilities results in the replacement of the functions f_{kl} by the functions $f_{kl} + g_k - g_l + b_{kl}$, and hence

c_{klm} by numbers

$$\hat{c}_{klm} = c_{klm} + (b_{kl} + b_{lm} + b_{mk}). \tag{30}$$

(Notice that the dependence on functions g_k disappears at the last step.)

Thus the form ω defines a skew-symmetric function c: $klm \mapsto c_{klm}$ up to equivalence described by relation (30).

We shall call skew-symmetric functions c: $klm \mapsto c_{klm}$ *two-dimensional cocycles* of octahedron (13) and call cocycles of the form

$$c_{klm} = b_{kl} + b_{lm} + b_{mk} \tag{31}$$

coboundaries. Cocycles form a vector space Z^2 (whose dimension is equal to the number of faces of octahedron (13), i.e. to 8) and the coboundaries form its subspace B^2. The corresponding quotient space Z^2/B^2 is designated as H^2. Its elements, i.e. cosets of Z^2 modulo the subspace B^2, are called *two-dimensional cohomology classes* of octahedron (13).

Since relations (30) precisely mean that cocycles c: $klm \mapsto c_{klm}$ and $\hat{c}$: $klm \mapsto \hat{c}_{klm}$ belong to the same cohomology class $[c]$, we see that we have constructed some (obviously linear) mapping

$$\Omega^2 S^2 \to H^2, \quad \omega \mapsto [c]. \tag{32}$$

If ω is an exact form, i.e. has the form $d\alpha$, where $\alpha \in \Omega^1 S^2$, then we may take as forms α_k restrictions $\alpha \,|_{U_k}$ of the form α on coordinate neighbourhoods U_k. Then differences $\alpha_k - \alpha_l$ are zero, and we can satisfy relations (29) by taking as f_{kl} identically zero functions. Since with this choice of functions f_{kl} all numbers c_{klm} are obviously zero, this proves that *under mapping* (32) *all exact forms change into the zero of the space* H^2 (the cohomology class [0] of a zero cocycle) and therefore proves that that mapping induces some (also linear) mapping

$$H^2 S^2 \to H^2. \tag{33}$$

By analogy with the case of $H^1 S^1$ it is natural to expect that *this mapping is a monomorphism*, i.e. if a cocycle $c\colon klm \mapsto c_{klm}$ corresponding to ω is of the form (31), then ω is an exact form. But if relation (31) is satisfied, then, replacing functions f_{kl} by differences $f_{kl} - b_{kl}$ we get $c_{klm} = 0$. Consequently, in this case we may assume without loss of generality that for any face klm of octahedron (13) there is a relation

$$(34) \qquad f_{kl} + f_{lm} + f_{mk} = 0 \text{ on } U_k \cap U_l \cap U_m.$$

Lemma 3. *If functions* $f_{kl} = -f_{lk}$ *satisfy relation* (34) *then there are functions* $g_k\colon U_k \to \mathbb{R}$ *such that*

$$f_{kl} = g_k - g_l \text{ on } U_k \cap U_l$$

for any edge kl of octahedron (13).

In Lecture 22 we shall prove the general Theorem 1 of which Lemma 3 is a special case. For the time being, however, we shall take this lemma for granted.

Problem 4. Relation (34) is identical in form with relation (15) of Lemma 2. Making use of this (and of Remark 1) prove Lemma 3.

It follows from Lemma 3 that forms $\beta_k = \alpha_k - dg_k$ satisfy relations

$$\beta_k = \beta_l \text{ on } U_k \cap U_l$$

for any edge kl of (13) and make up therefore some form β on $\mathbb{S}^2$. Since for any k

$$d\beta_k = d\alpha_k = \omega \text{ on } U_k,$$

we have $d\beta = \omega$ on $\mathbb{S}^2$, and hence ω is an exact form.

Thus mapping (33) is indeed a monomorphism. □

Further the computation proceeds along the already familiar lines. We first prove that $H^2 \simeq \mathbb{R}$ (with the aid of the functional Ind: $Z^2 \to \mathbb{R}$ equal to the sum of numbers c_{klm} over appropriately oriented faces of (13)) and then produce a form ω_0 for which the cocycle c has the property that Ind $c \neq 0$. Finally we get $H^2\mathbb{S}^2 \simeq \mathbb{R}$, i.e.

$$(35) \qquad h^2\mathbb{S}^2 = 1.$$

Problem 5. Carry out in detail the sketched proof of equation (35). [**Hint.** Take as ω_0 a form

$$\omega_0 = x\,dy \wedge dz - y\,dx \wedge dz + z\,dx \wedge dy.$$

To establish that this form is not exact is easiest if we first prove using the Stokes formula (we may get along with Green's formula as well) that the *integral of an exact form over the sphere* $\mathbb{S}^2$ *is equal to zero*, and then compute that the integral of the form ω_0 is equal to 4π.]

We see that even for the simplest manifolds computing cohomology groups runs into certain difficulties of a combinatorial and of an analytic nature. As to combinatorial difficulties, however, it is sometimes possible to successfully overcome or at least reduce them, by an appropriate choice of cover. For example, computing the group $H^1\mathbb{S}^2$ becomes significantly simpler if we make use of the two-element atlas $\{(U, h), (V, k)\}$ of Example 10 of Lecture 6. We obtain for this atlas a segment

rather than an octahedron and the proof of equation (18) becomes trivial. [According to the Poincaré lemma $\alpha = df_U$ on U and $\alpha = df_V$ on V, with $f_U - f_V = b$ on $U \cap V$, where b is some constant. Therefore the formula

$$f = \begin{cases} f_U & \text{on } U, \\ f_V + b & \text{on } V \end{cases}$$

correctly defines a function f on $\mathbb{S}^2$ such that $df = \alpha$.]

Since the last proof can be carried over word for word to the case of a sphere $\mathbb{S}^n$ of an arbitrary dimension $n \geqslant 2$ we even get

$$h^1\mathbb{S}^n = 0 \text{ for } n \geqslant 2. \tag{36}$$

If, however, we try to compute numbers $h^m\mathbb{S}^n$ for $m \geqslant 2$ by this method, then among other things we encounter the difficulty that the Poincaré lemma does not dirrectly apply to the intersection $U \cap V$ of charts U and V (since this intersection is diffeomorphic to a punctured space $\mathbb{R}^n \setminus \{0\}$ rather than to a ball). Nevertheless computation turns out to be possible if instead of the Poincaré lemma we make use of the general Proposition 3 and proceed by induction on n.

First consider the case where $0 < m < n$.

Proposition 4. *When $0 < m < n$ there is an equation*

$$h^m \mathbb{S}^n = 0. \tag{37}$$

Proof. For $n = 2$ (with $m = 1$) equation (37) is already known to us (see (18) or (36)). We proceed by induction on n. We may assume that $m \geqslant 2$, since for $m = 1$ equation (37) has been already proved above.

Suppose it has already been proved that $h^m \mathbb{S}^{n-1} = 0$ for $0 < m < n - 1$, and let ω be an arbitrary closed form of degree m on $\mathbb{S}^n$, where $0 < m < n$. Consider the two-element cover $\{U, V\}$ of a sphere $\mathbb{S}^n$ of Example 10 of Lecture 6.

Problem 6. Show that the *set* $U \cap V = \mathbb{S}^n \setminus \{p_0, q_0\}$ *is diffeomorphic to the product* $\mathbb{S}^{n-1} \times \mathring{\mathrm{I}}$. [**Hint.** Construct diffeomorphisms $U \cap V \to \mathbb{R}^n \setminus \{0\}$ and $\mathbb{S}^{n-1} \times \mathring{\mathrm{I}} \to \mathbb{R}^n \setminus \{0\}$.]

According to the Poincaré lemma there are forms θ_U and θ_V of degree $m - 1$ on U and V such that $\omega = d\theta_U$ on U and $\omega = d\theta_V$ on V. The form $\theta_U - \theta_V$ is closed on $U \cap V$ and is of degree $m - 1$. But according to the statement of Problem 6 and to Proposition 3 the Betti number $h^{m-1}\mathbb{S}^{n-1}$ of $U \cap V$ is equal to the Betti number $h^{m-1}\mathbb{S}^{n-1}$ of the sphere $\mathbb{S}^{n-1}$ and hence under the induction hypothesis is zero. There is therefore a form α of degree $m - 2$ on $U \cap V$ such that

$$\theta_U - \theta_V = d\alpha.$$

Problem 7. Construct on $\mathbb{S}^n$ a smooth function f assuming values in the interval $[0, 1]$ and equal to zero near a point q_0 (i.e. in some neighbourhood of that point) and to unity near a point p_0. [**Hint.** Cf. Corollary 2 to Lemma 1 of Lecture 1 and Proposition 2 of Lecture 14.]

Using the function f of Problem 7 we define a form α_U on U and a form α_V on V by putting

$$(\alpha_U)_p = \begin{cases} f(p)\,\alpha_p & \text{if } p \in U \cap V, \\ 0 & \text{otherwise (i.e. if } p = p_0), \end{cases}$$

$$(\alpha_V)_p = \begin{cases} (f(p) - 1)\,\alpha_p & \text{if } p \in U \cap V, \\ 0 & \text{otherwise (i.e. if } p = q_0). \end{cases}$$

Clearly forms α_U and α_V are smooth and we have on $U \cap V$ an equation $\alpha = \alpha_U - \alpha_V$ and hence an equation

$$\theta_U - d\alpha_U = \theta_V - d\alpha_V \text{ on } U \cap V.$$

Therefore the formula

$$\theta = \begin{cases} \theta_U - d\alpha_U \text{ on } U, \\ \theta_V - d\alpha_V \text{ on } V \end{cases}$$

correctly defines on the entire sphere S^n some form θ for which $d\theta = \omega$ on S^n. Since ω was an arbitrary closed form of degree m on S^n this proves that $h^m S^n = 0$.

This completes the proof of Proposition 4 by induction. □

The case $m = n$ can be treated in a similar way.

Proposition 5. *For any $n \geqslant 1$ there is an equation*

$$h^n S^n = 1. \tag{38}$$

Proof. Since for $n = 1$ equation (38) is already proved (see formula (26)) we can again make use of induction on n.

Let $\{U, V\}$ be the same cover of S^n as that above. Since under the induction hypothesis $h^{n-1} S^{n-1} = 1$ and hence

$$h^{n-1}(U \cap V) = h^{n-1}(S^{n-1} \times \mathring{I}) = h^{n-1} S^{n-1} = 1,$$

there is a closed but not exact form $\theta^{(0)}$ of degree $n - 1$ on $U \cap V$ having the property that any form of degree $n - 1$ on $U \cap V$ is cohomologous to a form $a\theta^{(0)}$, where a is some number.

Using the function f of Problem 7 we can represent the form $\theta^{(0)}$ by the already familiar method as

$$\theta^{(0)} = \theta_U^{(0)} - \theta_V^{(0)} \text{ on } U \cap V, \tag{39}$$

where $\theta_U^{(0)}$ and $\theta_V^{(0)}$ are some forms of degree $n - 1$ on U and V, respectively. Since under the hypothesis $d\theta^{(0)} = 0$, according to (39)

$$d\theta_U^{(0)} = d\theta_V^{(0)} \text{ on } U \cap V,$$

and therefore equations

$$\omega^{(0)}=\begin{cases} d\theta_U^{(0)} & \text{on } U,\\ d\theta_V^{(0)} & \text{on } V\end{cases}$$

correctly define on S^n some form $\omega^{(0)}$ of degree n.

Now let ω be an arbitrary form of degree n on S^n (trivially closed). According to the Poincaré lemma there are forms θ_U and θ_V of degree $n-1$ on U and V such that $\omega = d\theta_U$ on U and $\omega = d\theta_V$ on V. The form $\theta_U - \theta_V$ is closed on $U \cap V$, and there are therefore a form α of degree $n-2$ and a number $a \in \mathbb{R}$ on $U \cap V$ such that

$$\theta_U - \theta_V = d\alpha + a\theta^{(0)} \text{ on } U \cap V.$$

Representing as earlier the form α as $\alpha = \alpha_U - \alpha_V$, where α_U and α_V are forms on U and V, respectively, we can rewrite the equation as follows:

$$\theta_U - d\alpha_U - a\theta_U^{(0)} = \theta_V - d\alpha_V - a\theta_V^{(0)} \text{ on } U \cap V.$$

Therefore the formula

$$\beta=\begin{cases} \theta_U - d\alpha_U - a\theta_U^{(0)} & \text{on } U,\\ \theta_V - d\alpha_V - a\theta_V^{(0)} & \text{on } V\end{cases}$$

correctly defines on S^n some form β of degree $n-1$, with

$$d\beta = d\theta_U - a\, d\theta_U^{(0)} = \omega - a\omega^{(0)} \text{ on } U$$

and

$$d\beta = d\theta_V - a\, d\theta_U^{(0)} = \omega - a\omega^{(0)} \text{ on } V.$$

Hence $d\beta = \omega - a\omega^{(0)}$ on the entire sphere S^n, i.e. the form ω is cohomologous to $a\omega^{(0)}$.

To complete the proof of equation (38) it remains to be shown that the form $\omega^{(0)}$ is not cohomologous to zero (is not an exact form). But if $\omega^{(0)} = d\theta$, then there is an equation $d\theta_U^{(0)} = d(\theta\,|_U)$ on U and hence an equation $\theta_U^{(0)} = \theta\,|_U + d\varphi_U$, where φ_U is a form of degree $n-2$ on U. Similarly $\theta_V^{(0)} = \theta\,|_V + d\varphi_V$, where φ_V is a smooth function on V.

Hence on $U \cap V$

$$\theta^{(0)} = \theta_U^{(0)} - \theta_V^{(0)} = d(\varphi_U - \varphi_V),$$

i.e. the form $\theta^{(0)}$ is, contrary to the assumption, an exact form. The equation $\omega^{(0)} = d\theta$ is therefore impossible. □

Using the example of spheres S^1, S^2, and S^n this lecture has demonstrated the three major ways of computing cohomology groups of a manifold (with the aid of coverings in which intersections of any subfamilies are diffeomorphic to a ball, with the aid of more general coverings failing to satisfy this condition, and with the aid of integrals). In the lectures to follow we consider these methods more systematically. The example of spheres will serve as a pattern and a starting point although as a rule there will be no explicit references to this example.

Lecture 21

Simplicial schemes and their geometric realizations · Cohomology groups of simplicial schemes · The double complex of a covering · Cohomology groups of a double complex · Augmented double complexes · Edge homomorphisms · Acyclic complexes · Row acyclicity for $p = 0$.

We begin by describing in abstract form combinatorial schemes of mutual intersections of elements of coverings of smooth manifolds (or, more generally, of topological spaces).

Let $\mathscr{X}$ be an arbitrary topological space and $\mathfrak{U} = \{U_k, k \in K\}$ be its arbitrary open covering. A finite subset $\{k_0, \ldots, k_m\}$ of an index set K is said to be *distinguished* if the intersection $U_{k_0} \cap \ldots \cap U_{k_m}$ is not empty:

$$U_{k_0} \cap \ldots \cap U_{k_m} \neq \varnothing.$$

Definition 1. A set K in which some finite subsets are distinguished is said to be a *simplicial scheme* if any subset of a distinguished subset is distinguished.

It is clear that distinguished subsets of the index set K of $\mathfrak{U}$ satisfy this condition, and hence K is a simplicial scheme. This scheme is called the *nerve* of the covering $\mathfrak{U}$.

It is said that $m + 1$ points of an affine space are *affinely independent* if they are not contained in any $(m - 1)$-dimensional plane, i.e. if for their radius vectors $\mathbf{k}_0, \mathbf{k}_1, \ldots, \mathbf{k}_m$ vectors $\mathbf{k}_1 - \mathbf{k}_0, \ldots, \mathbf{k}_m - \mathbf{k}_0$ are linearly independent.

For any $m + 1$ affinely independent points $\mathbf{k}_0, \ldots, \mathbf{k}_m$ of $\mathbb{R}^n$ (or more generally, of an arbitrary vector space over the field $\mathbb{R}$) the set of all points $\mathbf{k}$ of the form

$$\mathbf{k} = t_0\mathbf{k}_0 + \ldots + t_m\mathbf{k}_m,$$

where $t_0, \ldots, t_m$ are real numbers such that

$$0 \leqslant t_0 \leqslant 1, \ldots, 0 \leqslant t_m \leqslant 1$$

and

$$t_0 + \ldots + t_m = 1$$

(we as ever identify points of $\mathbb{R}^n$ with their radius vectors) is called an *m-dimensional simplex with vertices* $\mathbf{k}_0, \ldots, \mathbf{k}_m$ and designated as $\mathbf{k}_0 \ldots \mathbf{k}_m$. This is a point $\mathbf{k}_0$ for $m = 0$, a segment $\mathbf{k}_0\mathbf{k}_1$ for $m = 1$, a triangle $\mathbf{k}_0\mathbf{k}_1\mathbf{k}_2$ for $m = 2$, and a tetrahedron $\mathbf{k}_0\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3$ for $m = 3$.

Numbers $t_0, \ldots, t_m$ for a point $\mathbf{k}$ of a simplex $\mathbf{k}_0 \ldots \mathbf{k}_m$ are called the *barycentric coordinates* of the point.

A point $\mathbf{k}$ of a simplex $\mathbf{k}_0 \ldots \mathbf{k}_m$ is said to be an *interior point* of the simplex if $0 < t_i < 1$ for all $i = 0, \ldots, m$. (Such a point is an interior point of $\mathbf{k}_0 \ldots \mathbf{k}_m$ in the general topological sense with respect to the m-dimensional plane containing the simplex.)

A simplicial scheme K is said to be *realizable* if its elements can be represented (with n sufficiently large) by points of $\mathbb{R}^n$ such that:

(i) for any distinguished set $\{k_0, \ldots, k_m\} \subset K$ the corresponding points $\mathbf{k}_0, \ldots, \mathbf{k}_m$ are affinely independent (and hence define a simplex $\mathbf{k}_0 \ldots \mathbf{k}_m$);

(ii) simplexes $\mathbf{k}_0 \ldots \mathbf{k}_m$ and $\mathbf{k}'_0 \ldots \mathbf{k}'_{m'}$ corresponding to two distinct distinguished subsets $\{k_0, \ldots, k_m\}$ and $\{k'_0, \ldots, k'_{m'}\}$ of K have no common interior points.

In this case the union of all simplexes $\mathbf{k}_0 \ldots \mathbf{k}_m$ corresponding to all possible distinguished subsets $\{k_0, \ldots, k_m\}$ is called a *geometric realization* of the scheme K. A geometric realization is designated $|K|$.

The pictures we drew in the preceding lecture were just the geometric realizations of the nerves of the corresponding coverings of the sphere $\mathbb{S}^2$ and of the circle $\mathbb{S}^1$.

According to a tendency in mathematics to transfer terms it is customary to call distinguished subsets of a scheme K its *simplexes*. [To distinguish when necessary simplexes in K from simplexes in $\mathbb{R}^n$, the former are called *abstract* (or *scheme*) simplexes and the latter *Eucli-*

dean (or *geometric*) simplexes.] Scheme elements involved in at least one simplex of the scheme are called its *vertices*, and a simplex $\{k_0, \ldots, k_m\}$ is called a *simplex with vertices* $k_0, \ldots, k_m$.

Elements of K that are not vertices will not be involved, in what follows, in any construction (just as in the geometric realization $|K|$) and can be safely ignored. [In defining schemes many authors introduce an additional axiom which requires that each element of a scheme should be its vertex. But then in defining the nerve of a covering it is necessary to require that the covering should] consist of nonempty sets, which is not always convenient.

Remark 1. Actually, we shall nowhere need geometric realizations. We have described them just in order to be able to use a geometric language with respect to simplicial schemes. Therefore we shall not, in particular, consider here a rather delicate question of what simplicial schemes are realizable. We just notice that any finite simplicial scheme is realizable in the obvious way.

Problem 1. A simplicial scheme K is said to be *realizable as an embedding* if there is a geometric realization $|K|$ of the scheme such that a subset $F \subset |K|$ is closed (in the induced topology) if and only if so is the intersection $F \cap \mathbf{k}_0 \ldots \mathbf{k}_m$ for any simplex $\{k_0, \ldots, k_m\} \subset K$. Show that a simplicial scheme is embedding-realizable if and only if:

(a) the set of its vertices is finite or countable;

(b) each of its vertices only belongs to a finite number of simplexes;

(c) there is an n such that every simplex of K contains at most $n + 1$ vertices.

Remark 2. Subsets of $\mathbb{R}^n$ that are geometric realizations of simplicial schemes are called *polyhedra* or *polytopes*. Their topological theory, which is generally called **piecewise linear topology**, is closely related to the topological theory of smooth manifolds (it can be shown—this is a difficult theorem!—that any smooth Hausdorff manifold with countable base is homeomorphic to some polytope) and highly advanced, constituting one of the most geometry-oriented parts of modern topology. Unfortunately, this theory is entirely beyond the scope of this course of lectures.

Let K be an arbitrary simplicial scheme, and let K_m, $m \geqslant 0$, be a subset of

$$\underbrace{K \times \ldots \times K}_{m+1 \text{ times}}$$

consisting of sequences $(k_0, \ldots, k_m)$, $k_0, \ldots, k_m \in K$, such that the set $\{k_0, \ldots, k_m\}$ is a simplex of K. (Elements of K_m are generally called *m-dimensional ordered simplexes* of a scheme K.)

Also let G be an arbitrary Abelian group (in additive notation).

Definition 2. A mapping

$$c\colon K_m \to G, \quad (k_0, \ldots, k_m) \mapsto c\,(k_0, \ldots, k_m)$$

is said to be an *m-dimensional cochain of a scheme K over G* (or with *values* in G) if this mapping is skew-symmetric, i.e. if $c\,(k_0, \ldots, k_m)$ changes sign when any two adjacent arguments k_a, k_{a+1}, $0 \leqslant a \leqslant m-1$, are interchanged. In other words, the mapping c is a cochain if

$$(1) \qquad c\,(k_{\sigma(0)}, \ldots, k_{\sigma(m)}) = \varepsilon_\sigma c\,(k_0, \ldots, k_m)$$

for any ordered simplex $(k_0, \ldots, k_m) \in K_m$ and any permutation σ of indices $0, 1, \ldots, m$ (where as ever ε_σ is the sign of permutation σ).

The set of all such cochains is obviously a group. This will be designated $C^m\,(K;\, G)$.

When G is a field the group $C^m\,(K;\, G)$ is a vector space over that field. In particular, for $G = \mathbb{R}$ (the only case of interest to us now!) $C^m(K;\, \mathbb{R})$ is a vector space over the field $\mathbb{R}$.

It can easily be seen that for any cochain $c \in C^m\,(K;\, G)$ the formula

$$(2) \qquad (\delta c)\,(k_0, \ldots, k_{m+1}) = \sum_{i=0}^{m+1} (-1)^i c\,(k_0, \ldots, \hat{k}_i, \ldots, k_{m+1}),$$

where ^ indicates that the corresponding element k_i must be omitted, defines some cochain $\delta c \in C^{m+1}\,(K;\, G)$. (Indeed, when two adjacent vertices k_a and k_{a+1}, $0 \leqslant a \leqslant m$, are interchanged all terms of sum (2) change sign, except for the ath and $(a+1)$st terms which, entering as they do into the sum with different signs, are interchanged.)

Clearly, the mapping

$$\delta\colon C^m(K;\,G) \to C^{m+1}(K;\,G),\ c \mapsto \delta c,$$

is a homomorphism. This is called a *coboundary operator*.

Since δ is defined for any $m \geqslant 0$, we can make up a sequence

$$C^{\cdot}(K;\,G)\colon C^0(K;\,G) \xrightarrow{\delta} C^1(K;\,G)$$
$$\to \ldots \to C^m(K;\,G) \xrightarrow{\delta} C^{m+1}(K;\,G) \to \ldots$$

Proposition 1. *A sequence $C^{\cdot}(K;\,G)$ is a cochain complex, i.e.*

$$\delta(\delta c) = 0$$

for every cochain $c \in C^m(K;\,G)$ and every $m \geqslant 0$.

Proof. For any ordered complex $(k_0, \ldots, k_{m+2}) \in K_{m+2}$ the element $\delta(\delta c)(k_0, \ldots, k_{m+2})$ of a group G is the sum of elements of the form

$$c(k_0, \ldots, \hat{k}_i, \ldots, \hat{k}_j, \ldots, k_{m+2}),\quad 0 \leqslant i \leqslant j \leqslant m+2, \tag{3}$$

each occurring twice: the first time in computing the term $(-1)^j(\delta c)(k_0, \ldots, \hat{k}_j, \ldots, k_{m+2})$ and the second time in computing the term $(-1)^i(\delta c)(k_0, \ldots, \hat{k}_i, \ldots, k_{m+2})$. In the former case the term (3) has a sign $(-1)^{i+j}$ (since in the simplex $(k_0, \ldots, \hat{k}_j, \ldots, k_{m+2})$ the vertex k_i has a number i) and in the latter case it has a sign $(-1)^{i+j-1}$ (since in the simplex $(k_0, \ldots, \hat{k}_i, \ldots, k_{m+2})$ the vertex k_j has a number $j-1$). Therefore all terms in (3) are cancelled. ☐

Cochains $c \in C^m(K;\,G)$ for which $\delta c = 0$ are called *cocycles* of K over a group G, and cochains of the form δc, where $c \in C^{m-1}(K;\,G)$, are called *coboundaries*. Cocycles make up a subgroup $Z^m(K;\,G)$ (the kernel of a homomorphism $\delta\colon C^m(K;\,G) \to C^{m+1}(K;\,G)$) and coboundaries make up a subgroup $B^m(K;\,G)$ (the image of a homomorphism $\delta\colon C^{m-1}(K;\,G) \to C^m(K;\,G)$; it being assumed by convention that for $m=0$ $B^0(K;\,G) = 0$).

Since $B^m(K; G) \subset Z^m(K; G)$, a quotient group

$$H^m(K; G) = Z^m(K; G)/B^m(K; G)$$

is defined.

This quotient group is called an *m-dimensional* (or *m*th) *cohomology group of K over G*, and its elements are called *cohomology classes.* (Compare with the general definitions at the beginning of Lecture 20.)

Remark 3. It can be shown (it is a difficult theorem!) that cohomology groups (and dually defined homology groups; see Lecture 29 below) of two schemes are isomorphic if geometric realizations of those schemes are homeomorphic. This allows the device of cohomology (and homology) groups of simplicial schemes to be applied to the study of topological properties of polytopes. The corresponding branch of mathematics is called **combinatorial topology.** Some fifty years ago it was entirely independent, but nowadays combinatorial topology has almost completely dissolved in a more general **algebraic topology.** [The term "combinatorial topology" is sometimes applied to the entire algebraic topology, but at present this no longer corresponds to the actual situation.]

When K is the nerve of a covering $\mathfrak{A} = \{U_k, k \in K\}$ of a smooth manifold $\mathscr{X}$ the construction of a cochain complex $C^{\cdot}(K; G)$ (designated in this case as $C^{\cdot}(\mathfrak{A}; G)$) may be generalized. Let for definiteness $G = \mathbb{R}$.

By a *cochain of dimension* $p \geqslant 0$ *of* $\mathfrak{A}$ *with values in functions* is meant a mapping c, defined on K_p, which associates every ordered simplex $(k_0, \ldots, k_p) \in K_p$ of the nerve K of $\mathfrak{A}$ with a smooth function

$$c(k_0, \ldots, k_p): U_{k_0} \cap \ldots \cap U_{k_p} \to \mathbb{R}$$

defined on a (nonempty) open set $U_{k_0} \cap \ldots \cap U_{k_p}$ and satisfying the skew-symmetry condition (1). Such cochains form a vector space which is designated $C^{p,0}(\mathfrak{A})$. The coboundary $\delta c \in C^{p+1,0}(\mathfrak{A})$ of a cochain $c \in C^{p,0}(\mathfrak{A})$ is defined by the same formula (2) (in which m is replaced by p), the only difference being that every term $c(k_0, \ldots, \hat{k}_i, \ldots, k_{p+1})$ of the right-hand side is assumed to be bounded on $U_{k_0} \cap \ldots \cap U_{k_{p+1}}$ (formula (2) makes no sense without this). Clearly, the relation $\delta(\delta c) = 0$ (together with the proof) remains valid in this case as well, i.e. *the family* $C^{\cdot,0}(\mathfrak{A}) = \{C^{p,0}(\mathfrak{A}),\ \delta\}$ *of vector spaces* $C^{p,0}(\mathfrak{A})$ *and homomorphisms*

$$\delta\colon C^{p,0}(\mathfrak{A}) \to C^{p+1,0}(\mathfrak{A})$$

is a cochain complex.

More generally, for any $q \geqslant 0$ we can introduce into consideration a cochain complex

$$C^{\cdot,q}(\mathfrak{A}) = \{C^{p+q}(\mathfrak{A}),\ \delta\}$$

consisting of *cochains with values in forms of degree* q, i.e. of mapping c, defined on K_p, that satisfy the skew-symmetry condition (1) whose values $c(k_0, \ldots, k_p)$ are forms of degree q defined on $U_{k_0} \cap \ldots \cap U_{k_p}$ (elements of a vector space $\Omega^q(U_{k_0} \cap \ldots \cap U_{k_p})$). In this case the operator δ is defined by the same formula (2) (with the same refinements concerning the boundedness condition). For $q = 0$ we obtain the same complex $C^{\cdot,0}(\mathfrak{A})$.

Any cochain $c \in C^{p,q}(\mathfrak{A})$ can be associated with a cochain $dc \in C^{p,q+1}(\mathfrak{A})$ assuming by definition that

$$(dc)(k_0, \ldots, k_p) = dc(k_0, \ldots, k_p)$$

for any ordered simplex $(k_0, \ldots, k_p) \in K_p$.

Mappings

$$d\colon C^{p,q}(\mathfrak{A}) \to C^{p,q+1}(\mathfrak{A}) \tag{4}$$

clearly commute with operators δ, i.e. for any $p,\ q \geqslant 0$ the diagram

$$\begin{tikzcd}
C^{p,q}(\mathfrak{A}) \arrow[r, "\delta"] \arrow[d, "d"'] & C^{p+1,q}(\mathfrak{A}) \arrow[d, "d"'] \\
C^{p,q+1}(\mathfrak{A}) \arrow[r, "\delta"] & C^{p+1,q+1}(\mathfrak{A})
\end{tikzcd}$$

is commutative. This means by definition (see Definition 2 of Lecture 20) that mappings (4) make up a cochain mapping

$$d\colon C^{\cdot,q}(\mathfrak{A}) \to C^{\cdot,q+1}(\mathfrak{A})$$

of a complex $C^{\cdot,q}(\mathfrak{A})$ into a complex $C^{\cdot,q+1}(\mathfrak{A})$.

Notice that $d \circ d = 0$, i.e. for any $p \geqslant 0$ the family $C^{p,\cdot}(\mathfrak{A}) = \{C^{p,q}(\mathfrak{A}),\ d\}$ is a cochain complex, with mappings

$$\delta\colon C^{p,q}(\mathfrak{A}) \to C^{p+1,q}(\mathfrak{A})$$

making up a cochain mapping of a complex $C^{p,\cdot}(\mathfrak{A})$ into a complex $C^{p+1,\cdot}(\mathfrak{A})$.

Definition 3. A family $C^{\cdot,\cdot} = \{C^{p,q}, \delta, d\}$ of groups $C^{p,q}$, where $p, q \geqslant 0$ and of mappings

$$\delta\colon C^{p,q} \to C^{p+1,q}, \quad d\colon C^{p,q} \to C^{p,q+1}$$

is said to be a *double complex* if

$$\delta \circ \delta = 0, \; d \circ d = 0 \text{ and } d \circ \delta = \delta \circ d,$$

i.e. if

1° for any $q \geqslant 0$ (for any $p \geqslant 0$) the family $C^{\cdot,q} = \{C^{p,q}, \delta\}$ (the family $C^{p,\cdot} = \{C^{p,q}, d\}$) is a cochain complex;

2° mappings d make up a cochain mapping

$$d\colon C^{\cdot,q} \to C^{\cdot,q+1}$$

of a complex $C^{\cdot,q}$ into a complex $C^{\cdot,q+1}$, and hence mappings δ make up a cochain mapping

$$\delta\colon C^{p,\cdot} \to C^{p+1,\cdot}$$

of a complex $C^{p,\cdot}$ into a complex $C^{p+1,\cdot}$.

Thus we see that for any covering $\mathfrak{A}$ of a smooth manifold $\mathscr{X}$ we have constructed a double complex

$$C^{\cdot,\cdot}(\mathfrak{A}) = \{C^{p,q}(\mathfrak{A}), \delta, d\}.$$

We shall call this a *double complex of a covering* $\mathfrak{A}$.

Double complexes will be represented as arrays of the form

(5)

$$\begin{tikzcd}
\vdots & \vdots & & \vdots & \vdots & \\
C^{0,q+1} \arrow[u] \arrow[r] & C^{1,q+1} \arrow[u] \arrow[r] & \cdots \arrow[r] & C^{p,q+1} \arrow[u] \arrow[r, "\delta"] & C^{p+1,q+1} \arrow[u] \arrow[r] & \cdots \\
C^{0,q} \arrow[u] \arrow[r] & C^{1,0} \arrow[u] \arrow[r] & \cdots \arrow[r] & C^{p,q} \arrow[u, "a"] \arrow[r, "\delta"] & C^{p+1,q} \arrow[u, "d"'] \arrow[r] & \cdots \\
\vdots \arrow[u] & \vdots \arrow[u] & & \vdots \arrow[u] & \vdots \arrow[u] & \\
C^{0,0} \arrow[u] \arrow[r] & C^{1,0} \arrow[u] \arrow[r] & \cdots \arrow[r] & C^{p,0} \arrow[u] \arrow[r] & C^{p+1,0} \arrow[u] \arrow[r] & \cdots
\end{tikzcd}$$

Accordingly, we shall call complexes $C^{\cdot,q}$ *rows* of a double complex $C^{\cdot,\cdot}$ and complexes $C^{p,\cdot}$ *columns* of $C^{\cdot,\cdot}$.

Putting for any $m \geqslant 0$

$$C^m = C^{0,m} \oplus C^{1,m-1} \oplus \ldots \oplus C^{m,0}$$

we define a mapping

$$\mathfrak{d} : C^m \to C^{m+1}$$

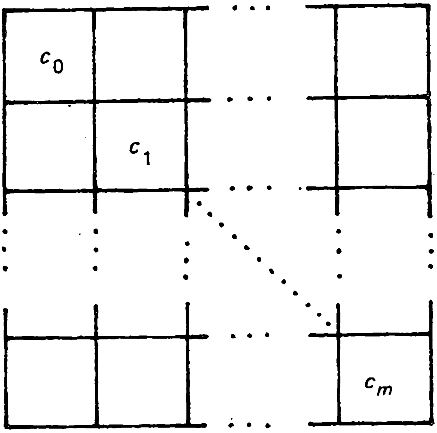
as follows:

$$\mathfrak{d}\, c = \delta c + d'c,$$

where

$$d'c = (-1)^p\, dc \text{ if } c \in C^{p,m-p}.$$

Elements of a group C^m are thus chains $c = (c_0, \ldots, c_m)$ of the cochains of a complex $C^{\cdot,\cdot}$ which lie on its mth antidiagonal:



and the mapping $\mathfrak{d}$ transforms every such chain into a chain $\mathfrak{d}c = (e_0, \ldots, e_{m+1})$ for which

$$\begin{aligned} e_0 &= d'c_0, \\ e_1 &= \delta c_0 + d'c_1, \\ &\ldots\ldots\ldots\ldots \\ e_m &= \delta c_{m-1} + d'c_m, \\ e_{m+1} &= \delta c_m, \end{aligned} \tag{6}$$

$$\begin{tikzcd}
e_0 & & & & \\
c_0 \arrow[u] \arrow[r] & e_1 & & & \\
& c_1 \arrow[u] \arrow[r] & \cdots & & \\
& & \vdots & & \\
& & c_{m-1} \arrow[u] \arrow[r] & e_m & \\
& & & c_m \arrow[u] \arrow[r] & e_{m+1}.
\end{tikzcd}$$

In particular, $\mathfrak{d}c = 0$ if and only if

$$
\begin{aligned}
d'c_0 &= 0,\\
\delta c_0 + d'c_1 &= 0,\\
&\cdots\cdots\cdots\cdots\\
\delta c_{m-1} + d'c_m &= 0,\\
\delta c_m &= 0,
\end{aligned}
\tag{7}
$$

$$\begin{tikzcd}
0 & & & & \\
c_0 \arrow[u] \arrow[r] & 0 & & & \\
& c_1 \arrow[u] \arrow[r] & \cdots & & \\
& & \vdots & & \\
& & c_{m-1} \arrow[u] \arrow[r] & 0 & \\
& & & c_m \arrow[u] \arrow[r] & 0.
\end{tikzcd}$$

Since $d'\delta c = (-1)^{p+1} d\delta c$ and $\delta d'c = (-1)^p \delta dc$, we have $d' \circ \delta + \delta \circ d' = 0$ and therefore

$$\mathfrak{d} \circ \mathfrak{d} = (\delta + d') \circ (\delta + d') = \delta \circ \delta + \delta \circ d' + d' \circ \delta + d' \circ d' = 0.$$

This means that the *family* $\{C^m, \mathfrak{d}\}$ *is a cochain complex.*

Cohomology groups of this cochain complex are called *cohomology groups of a double complex* $C^{\cdot,\cdot}$ and designated as $H^m(C^{\cdot,\cdot})$.

In the special case $C^{\cdot,\cdot} = C^{\cdot,\cdot}(\mathfrak{A})$ these groups are called *cohomology groups of a covering* $\mathfrak{A}$ and designated as $H^m(\mathfrak{A})$.

A double complex of $\mathfrak{A} = \{U_k, k \in K\}$ can be augmented in a natural way by a de Rham complex $\{\Omega^q\mathscr{X}, d\}$ of a manifold $\mathscr{X}$ at the left and a complex $\{C^p(\mathfrak{A}; \mathbb{R}), \delta\}$ of the cochains of $\mathfrak{A}$ with coefficients in the field $\mathbb{R}$ at the bottom:

(8)

$$\begin{tikzcd}
 & \vdots & \vdots & & & \vdots & \vdots & \\
0 \arrow[r] & \Omega^q \mathscr{X} \arrow[r, "j"] \arrow[u] & C^{0,q}(\mathfrak{U}) \arrow[r] \arrow[u] & \cdots \arrow[rr] & & C^{p,q}(\mathfrak{U}) \arrow[r, "\delta"'] \arrow[u, "d"'] & C^{p+1,q}(\mathfrak{U}) \arrow[r] \arrow[u, "d"] & \cdots \\
 & \vdots \arrow[u] & \vdots \arrow[u] & \vdots & & \vdots \arrow[u] & \vdots \arrow[u] & \\
0 \arrow[r] & \Omega^1 \mathscr{X} \arrow[r, "j"] \arrow[u] & C^{0,1}(\mathfrak{U}) \arrow[r] \arrow[u] & C^{1,1}(\mathfrak{U}) \arrow[r] \arrow[u] & \cdots & \vdots & \vdots & \\
0 \arrow[r] & \Omega^0 \mathscr{X} \arrow[r, "j"] \arrow[u] & C^{0,0}(\mathfrak{U}) \arrow[r] \arrow[u] & C^{1,0}(\mathfrak{U}) \arrow[r] \arrow[u] & \cdots \arrow[r] & C^{p,0}(\mathfrak{U}) \arrow[r] \arrow[u] & C^{p+1,0}(\mathfrak{U}) \arrow[r] \arrow[u] & \cdots \\
 & & C^0(\mathfrak{U}; \mathbb{R}) \arrow[r] \arrow[u, "i"] & C^1(\mathfrak{U}; \mathbb{R}) \arrow[r] \arrow[u, "i"] & \cdots \arrow[r] & C^p(\mathfrak{U}; \mathbb{R}) \arrow[r] \arrow[u, "i"] & C^{p+1}(\mathfrak{U}; \mathbb{R}) \arrow[r] \arrow[u, "i"] & \cdots \\
 & & 0 \arrow[u] & 0 \arrow[u] & & 0 \arrow[u] & 0 \arrow[u] &
\end{tikzcd}$$

where i are embeddings resulting from identifying an arbitrary number $\lambda \in \mathbb{R}$ with a constant function assuming a value λ, and mappings j are defined as follows:

$$(j\omega)(k) = \omega \mid_{U_k},$$

where $\omega \in \Omega^q \mathscr{X}$ and $k \in K$.

Mappings i and j clearly commute with operators δ and d, respectively, i.e. make up cochain mappings

$$i\colon C^{\cdot}(\mathfrak{U}; \mathbb{R}) \to C^{\cdot,0}(\mathfrak{U}), \quad j\colon \Omega^{\cdot} \mathscr{X} \to C^{0,\cdot}(\mathfrak{U}).$$

Besides, they are monomorphisms and satisfy the relations

$$d \circ i = 0, \quad \delta \circ j = 0.$$

(The relation $d \circ i = 0$ follows from the fact that if $f = \text{const}$, then $df = 0$ and the relation $\delta \circ j = 0$ follows from the fact that by definition

$$(\delta c)(k_0, k_1) = c(k_1) - c(k_0) \text{ on } U_{k_0} \cap U_{k_1}$$

for any ordered simplex $(k_0, k_1) \in K_1$ and hence if $c(k) = \omega$ on U_k, then $(\delta c)(k_0, k_1) = 0$.)

This situation is worthy of a special definition.

Definition 4. By an *augmented double complex* is meant an array of groups and homomorphisms of the form

(9)

$$\begin{tikzcd}
 & \vdots & \vdots & & & \vdots & \vdots & \\
C \arrow[r] & B^q \arrow[u, "d"] \arrow[r, "j"] & C^{0,q} \arrow[u] \arrow[rr] & & \cdots \arrow[r] & C^{p,q} \arrow[u, "d"] \arrow[r, "\delta"] & C^{p+1,q} \arrow[u] \arrow[r] & \cdots \\
 & \vdots \arrow[u] & \vdots \arrow[u] & \vdots & & \vdots \arrow[u] & \vdots \arrow[u] & \\
0 \arrow[r] & B^1 \arrow[u] \arrow[r, "j"] & C^{0,1} \arrow[u] \arrow[r] & C^{1,1} \arrow[u] \arrow[r] & \cdots \arrow[r] & C^{p,1} \arrow[u] \arrow[r] & C^{p+1,1} \arrow[u] \arrow[r] & \cdots \\
0 \arrow[r] & B^0 \arrow[u] \arrow[r, "j"] & C^{0,0} \arrow[u] \arrow[r] & C^{1,0} \arrow[u] \arrow[r] & \cdots \arrow[r] & C^{p,0} \arrow[u] \arrow[r] & C^{p+1,0} \arrow[u] \arrow[r] & \cdots \\
 & & A^0 \arrow[u, "i"] \arrow[r] & A^1 \arrow[u, "i"] \arrow[r] & \cdots \arrow[r] & A^p \arrow[u, "i"] \arrow[r, "\delta"] & A^{p+1} \arrow[u, "i"] \arrow[r] & \cdots \\
 & & 0 \arrow[u] & 0 \arrow[u] & & 0 \arrow[u] & 0 \arrow[u] &
\end{tikzcd}$$

where the portion inside the frame is a double complex in the sense of Definition 3 and where

1° the families $A^\cdot = \{A^p, \delta\}$ and $B^\cdot = \{B^q, d\}$ are cochain complexes;

2° the mappings i make up a cochain mapping of a complex $A^\cdot$ into a complex $C^{\cdot,0}$, which has the property that

(10) $$d \circ i = 0;$$

3° the mappings j make up a cochain mapping of a complex $B^\cdot$ into a complex $C^{0,\cdot}$, which has the property that

$$\delta \circ j = 0;$$

4° all the mappings i and j are monomorphisms.

It follows directly from (10) that identifying for any element $a \in A$ an element $ia \in C^{m,0}$ with a chain $(0, \ldots, 0, ia) \in C^m$ yields a cochain mapping i of a complex $A^\cdot$ into a complex $\{C^m, \mathfrak{d}\}$. Indeed, since $(d \circ i)\, a = 0$, we have $\mathfrak{d}(ia) = (0, \ldots, 0\ (\delta \circ i)\, a) = (0, \ldots, 0, i(\delta a)) =$

$i\,(\delta a)$. Homomorphisms

$$i^*\colon H^m(A^\cdot) \to H^m(C^{\cdot,\cdot}) \tag{11}$$

induced by this cochain mapping are called *lower edge homomorphisms.*

Similarly defined are *left edge homomorphisms*

$$j^*\colon H^m(B^\cdot) \to H^m(C^{\cdot,\cdot}). \tag{12}$$

For a double complex $C^{\cdot,\cdot}(\mathfrak{A})$ of a covering $\mathfrak{A}$ of a manifold $\mathscr{X}$ lower edge homomorphisms are of the form

$$i^*\colon H^m(\mathfrak{A};\, \mathbb{R}) \to H^m(\mathfrak{A})$$

and left ones are of the form

$$j^*\colon H^m\mathscr{X} \to H^m(\mathfrak{A}).$$

We shall say that an augmented double complex (9) is *acyclic in columns* if each of its columns is an exact sequence, i.e. for any p, $q \geqslant 0$ there is an equation

$$\operatorname{Ker}(d\colon C^{p,q} \to C^{p,q+1}) = \begin{cases} \operatorname{Im}(i\colon A^p \to C^{p,0}) & \text{if } q = 0, \\ \operatorname{Im}(d\colon C^{p,q-1} \to C^{p,q}) & \text{if } q > 0. \end{cases}$$

[For $q = 0$ this equation means that the monomorphism i isomorphically maps a group A^p onto a group $H^0(C^{p,\cdot}) = \operatorname{Ker}(d\colon C^{p,0} \to C^{p,1})$ of a complex $C^{p,\cdot}$ and for $q > 0$ this means that the cohomology group $H^q(C^{p,\cdot})$ of that complex is equal to zero.]

Similarly defined are double complexes *acyclic in rows* for which

$$\operatorname{Ker}(\delta\colon C^{p,q} \to C^{p+1,q}) = \begin{cases} \operatorname{Im}(j\colon B^q \to C^{0,q}) & \text{if } p = 0, \\ \operatorname{Im}(\delta\colon C^{p-1,q} \to C^{p,q}) & \text{if } p > 0. \end{cases}$$

A key to the computation of de Rham cohomology groups of smooth manifolds is the following algebraic lemma:

Lemma 1. *If an augmented double complex* (9) *is acyclic in columns* (*in rows*), *then for any* $m \geqslant 0$ *the lower edge homomorphism* (11) (*the left edge homomorphism* (12)) *is an isomorphism.*

Proof. Suppose for a cocycle $a \in A^m$ of a complex $A^\bullet$ there is an equation $i^*[a] = 0$, i.e. there is a chain $c = (c_0, \ldots, c_{m-1}) \in C^{m-1}$ such that $\mathfrak{d}c = ia$. By definition this means that

$$\begin{aligned} d'c_0 &= 0, \\ \delta c_0 + d'c_1 &= 0, \\ &\ldots\ldots\ldots \\ \delta c_{m-2} + d'c_{m-1} &= 0, \\ \delta c_{m-1} &= ia, \end{aligned} \tag{13}$$

$$\begin{tikzcd}
0 & & & & \\
c_0 \arrow[u] \arrow[r] & 0 & & & \\
& c_1 \arrow[u] \arrow[r] & {} & & \\
& & \cdots & & \\
& & c_{m-2} \arrow[u] \arrow[r] & 0 & \\
& & & c_{m-1} \arrow[u] \arrow[r] & ia.
\end{tikzcd}$$

Since under the hypothesis the first column of a complex $C^{\bullet,\bullet}$ is an exact sequence, there is a cochain $b_0 \in C^{0,m-1}$ such that $db_0 = c_0$. Then the cochain $c' = c - \mathfrak{d}b_0$ (more exactly, the cochain $c - \mathfrak{d}(b_0, 0, \ldots, 0)$) is of the form $(0, c'_1, \ldots, c'_{m-1})$ and as before satisfies the relation $\mathfrak{d}c' = ia$. Therefore we may assume without loss of generality from the onset that $c_0 = 0$.

But then the second relation of (13) will be of the form $dc_1 = 0$, and since the second column of $C^{\bullet,\bullet}$ is exact, there will therefore be an element $b_1 \in C^{1,m-1}$ such that $db_1 = c_1$. Replacing c by $c - \mathfrak{d}b_1$ we therefore get $c_0 = 0$ and $c_1 = 0$ (preserving the relation $\mathfrak{d}c = ia$).

Moving in this way along the antidiagonal downwards we finally get a chain c of the form $(0, \ldots, 0, c_{m-1})$, where c_{m-1} is a cochain of $C^{m-1,0}$ such that $dc_{m-1} = 0$ and $\delta c_{m-1} = ia$. But since $dc_{m-1} = 0$, again in view of the corresponding column being exact, there is an element $a' \in A^{m-1}$ such that $ia' = c_{m-1}$ and hence such that

$$i(\delta a') = \delta(ia') = \delta c_{m-1} = ia.$$

Since under the hypothesis the mapping i is monomorphic, this proves that $\delta a' = a$, i.e. $[a] = 0$ in $H^m(A^\bullet)$. Consequently, homomorphism (11) is a monomorphism.

Now let $c = (c_0, \ldots, c_{m-1})$ be an arbitrary m-dimensional cocycle of a complex $\{C^m, \mathfrak{d}\}$. Then relations (7) hold for cochains $c_0, \ldots, c_m$. In particular, $dc_0 = 0$. Again there is therefore a cochain b_0 such that for a cocycle $c - \mathfrak{d}b_0$ the component c_0 is zero, and hence, by virtue of the second relation (7), there is for this cocycle an equation $dc_1 = 0$ and, consequently, there is a cochain b_1 such that in the cocycle $c - db_0 - db_1$ the component c_1 is also zero.

Continuing the process we ultimately get a cocycle cohomologous to the given one all of whose components $c_0, \ldots, c_{m-1}$ are zero and the last component c_m is of the form ia, where a is some cocycle of a complex $A^\bullet$, i.e. which is the image of a cocycle a under a mapping $i\colon A^m \to C^m$. This means that homomorphism (11) is an epimorphism.

A symmetric statement about double complexes acyclic in rows can be proved in a similar way. □

Corollary 1. *If an augmented double complex* (9) *is acyclic in rows and in columns, then for any* $m \geqslant 0$ *the group* $H^m(A^\bullet)$ *is isomorphic to the group* $H^m(B^\bullet)$. □

Corollary 2. *If for an open covering* $\mathfrak{A} = \{U_k, k \in K\}$ *of a smooth manifold* $\mathscr{X}$ *the augmented complex* (8) *is acyclic in rows and in columns, then for any* $m \geqslant 0$ *the de Rham cohomology group* $H^m\mathscr{X}$ *of a manifold* $\mathscr{X}$ *is isomorphic to the cohomology group* $H^m(\mathfrak{A}; \mathbb{R})$ *of a covering* $\mathfrak{A}$:

$$H^m\mathscr{X} \simeq H^m(\mathfrak{A}; \mathbb{R}). \ \square$$

In this case we can thus compute the group $H^m\mathscr{X}$ in a purely combinatorial way from the nerve K of $\mathfrak{A}$. Compare with computing the group $H^m\mathbb{S}^n$ in the preceding lecture.)

The acyclicity condition for rows of the double complex (8) for $p = 0$ means that every cocycle $c \in C^{0,q}(\mathfrak{A})$ is of the form $i\omega$, where $\omega \in \Omega^q\mathscr{X}$, i.e. $c(k) = \omega$ on U_k for any $k \in K$. But this is really the case since the condition that the cochain $c\colon k \mapsto c(k) \in \Omega^q U_k$ should be a cocycle implies that

$$c(k_0) = c(k_1) \quad \text{on} \quad U_{k_0} \cap U_{k_1}$$

for any two indices k_0, $k_1 \in K$ (for which $U_{k_0} \cap U_{k_1} \neq \varnothing$), i.e. forms $c(k)$, $k \in K$, are compatible on intersections and hence make up some form $\omega \in \Omega^q\mathscr{X}$. Thus the *acyclicity condition for rows holds, with* $p = 0$, *for complex* (8) *of an arbitrary covering* $\mathfrak{A}$.

The conditions ensuring that rows and columns of (8) are acyclic (for $p > 0$ and $q \geqslant 0$, respectively) are considered in the next lecture.

Lecture 22

Row acyclicity of the double complex of a numerable covering · Row acyclicity of the double complex of a Leray covering · The de Rham-Leray theorem · Generalization · Groups $E_2^{p,q}$ · *Groups* $F^{p,q}$ · *A group adjoined to a graded group with filtration*

Definition 1. A family $\{\eta\}$ of smooth nonnegative functions η_k: $\mathscr{X} \to \mathbb{R}$ is said to be *locally finite* if for any point of a manifold $\mathscr{X}$ there is a neighbourhood of the point in which only a finite number of functions η_k are nonzero. A locally finite family $\{\eta_k\}$ of functions is said to be a *partition of unity* if

$$\sum_k \eta_k = 1$$

(this sum makes sense due to the local finiteness condition). A partition of unity $\{\eta_k\}$ is said to be *subordinate* to an open covering $\{U_k\}$ (with the same index set) if $\eta_k = 0$ outside U_k for any k. (Notice that there is another, more restricted definition in the literature, one that requires that U_k should contain not only a set where $\eta_k \neq 0$ but also its closure.) An open covering $\{U_k\}$ is said to be *numerable* if there is a subordinate partition of unity $\{\eta_k\}$. A manifold $\mathscr{X}$ is said to be *paracompact* if each open covering of it is numerable.

It turns out that *if a covering* $\mathfrak{A} = \{U_k;\ k \in K\}$ *of a smooth manifold* $\mathscr{X}$ *is numerable, then its double complex* $C^{\bullet,\bullet}(\mathfrak{A})$ *is acyclic in rows.* Indeed, according to the remark made at the end of the preceding lecture, it is only necessary for us to prove that for any $q \geqslant 0$ every

p-dimensional ($p > 0$) cocycle c of a cochain complex $C^{\cdot,q} = \{C^{p,q}(\mathfrak{A}), \delta\}$ is a coboundary. To this end we associate an arbitrary cochain $c \in C^{p,q}(\mathfrak{A})$ with a cochain $Dc \in C^{p-1,q}(\mathfrak{A})$ defined (obviously correctly) by the formula

$$(1)\qquad (Dc)(k_0, \ldots, k_{p-1}) = \sum_{k \in K} \eta_k c(k, k_0, \ldots, k_{p-1}),$$
$$(k_0, \ldots, k_{p-1}) \in K_{p-1},$$

where $\{\eta_k\}$ is a partition of unity subordinate to a covering $\mathfrak{A}$. Then for any simplex $(k_0, \ldots, k_p) \in K_p$

$$\begin{aligned}(\delta Dc)&(k_0, \ldots, k_p)\\ &= \sum_{i=0}^{p} (-1)^i \sum_{k \in K} \eta_k c(k, k_0, \ldots, \hat{k}_i, \ldots, k_p)\\ &= \sum_{k \in K} \eta_k \sum_{i=0}^{p} (-1)^i c(k, k_0, \ldots, \hat{k}_i, \ldots, k_p)\end{aligned}$$

and

$$\begin{aligned}(D\delta c)(k_0, \ldots, k_p) &= \sum_{k \in K} \eta_k (\delta c)(k, k_0, \ldots, k_p)\\ = \sum_{k \in K} \eta_k \Big[c(k_0, \ldots, k_p) &+ \sum_{i=0}^{p} (-1)^{i+1} c(k, k_0, \ldots, \hat{k}_i, \ldots, k_p)\Big]\\ = c(k_0, \ldots, k_p) &- \sum_{k \in K} \eta_k \sum_{i=0}^{p} (-1)^i c(k, k_0, \ldots, \hat{k}_i, \ldots, k_p)\\ = c(k_0, \ldots, k_p) &- (\delta Dc)(k_0, \ldots, k_p),\end{aligned}$$

i.e.

$$D\delta c + \delta Dc = c$$

for any cochain c. Hence if $\delta c = 0$, then $c = \delta Dc$. □

Now consider conditions ensuring that the complex $C^{\cdot,\cdot}(\mathfrak{A})$ is acyclic in columns.

Recall (cf. Definition 1 of Lecture 10) that for an arbitrary family of sets $\{\mathscr{X}_\alpha\}$ by $\prod\limits_\alpha \mathscr{X}_\alpha$ we denote the set of all possible families $\{x_\alpha\}$, where $x_\alpha \in \mathscr{X}_\alpha$. When all $\mathscr{X}_\alpha$ are vector spaces that set is a vector space with respect to componentwise operations ($\{x_\alpha\} + \{y_\alpha\} = \{x_\alpha + y_\alpha\}$, $\lambda\{x_\alpha\} = \{\lambda x_\alpha\}$) and called a *direct product* of vector spaces $\mathscr{X}_\alpha$. Similarly defined are direct products of groups, cochain complexes, double complexes, etc.

It is clear that, for example, *cohomology groups of a direct product of complexes are direct products of cohomology groups of factors.*

A comparison of the general definition of a direct product of vector spaces with the definition of vector spaces $C^{p,q}(\mathfrak{A})$ shows immediately that *every vector space $C^{p,q}(\mathfrak{A})$ is nothing but a direct product*

$$\prod_{\{k_0, \ldots, k_p\} \subset K} \Omega^q (U_{k_0} \cap \ldots \cap U_{k_p})$$

of vector spaces $\Omega^q(U_{k_0} \cap \ldots \cap U_{k_p})$ over all p-dimensional simplexes $\{k_0, \ldots, k_p\}$ of a simplicial scheme K. The mapping $d\colon C^{p,q}(\mathfrak{A}) \to C^{p,q+1}(\mathfrak{A})$ carries each factor $\Omega^q(U_{k_0} \cap \ldots \cap U_{k_p})$ of a vector space $C^{p,q}(\mathfrak{A})$ into a factor $\Omega^{q+1}(U_{k_0} \cap \ldots \cap U_{k_p})$ of $C^{p,q+1}(\mathfrak{A})$ and is nothing but an exterior differential

$$d\colon \Omega^q (U_{k_0} \cap \ldots \cap U_{k_p}) \to \Omega^{q+1} (U_{k_0} \cap \ldots \cap U_{k_p}).$$

This means that for any $p \geqslant 0$ the complex $C^{p,\cdot}(\mathfrak{A})$ is a direct product of de Rham complexes $\Omega^\cdot(U_{k_0} \cap \ldots \cap U_{k_p})$ and hence its cohomology groups are direct products of groups $H^q(U_{k_0} \cap \ldots \cap U_{k_p})$. Therefore for the complex $C^{\cdot,\cdot}(\mathfrak{A})$ the row acyclicity condition for $q > 0$ holds if and only if for any simplex $\{k_0, \ldots, k_p\} \subset K$ all cohomology groups $H^q(U_{k_0} \cap \ldots \cap U_{k_p})$, $q > 0$, are zero. Similarly, since for any $p \geqslant 0$ the group $C^p(\mathfrak{A}; \mathbb{R})$ is nothing but a direct product of vector spaces $\mathbb{R}$ over all simplexes $\{k_0, \ldots, k_p\} \subset K$, the column acyclicity condition for the complex $C^{\cdot,\cdot}(\mathfrak{A})$, with $q = 0$, is equivalent to the requirement that for any simplex $\{k_0, \ldots, k_p\} \subset K$ the group $H^0(U_{k_0} \cap \ldots \cap U_{k_p})$ should be isomorphic to $\mathbb{R}$.

Since by virtue of the Poincaré lemma (and Proposition 2 of Lecture 20) both conditions are trivially satisfied when all intersections $U_{k_0} \cap \ldots \cap U_{k_p}$ are diffeomorphic to an open ball $\overset{\circ}{\mathbb{B}}{}^n$ of $\mathbb{R}^n$ (where as ever $n = \dim \mathscr{X}$), we see that the *complex* $C^{\cdot,\cdot}(\mathfrak{A})$ *is acyclic in columns if for any indices* $k_0, \ldots, k_p \in K$ *the intersection* $U_{k_0} \cap \ldots \cap U_{k_p}$, *if not empty, is diffeomorphic to the ball* $\overset{\circ}{\mathbb{B}}{}^n$.

Definition 2. An open covering $\mathfrak{A} = \{U_k;\ k \in K\}$ of a smooth n-dimensional manifold $\mathscr{X}$ is said to be a *Leray covering* if for any indices $k_0, \ldots, k_p \in K$ the intersection $U_{k_0} \cap \ldots \cap U_{k_p}$ is either empty or diffeomorphic to a ball $\overset{\circ}{\mathbb{B}}{}^n$.

Thus, *for any Leray covering* $\mathfrak{A}$ *the complex* $C^{\cdot,\cdot}(\mathfrak{A})$ *is acyclic in columns.*

On comparing the results obtained we immediately get the following theorem by virtue of Corollary 2 to Lemma 1 of the preceding lecture.

Theorem 1. *For any numerable Leray covering of a smooth manifold* $\mathscr{X}$ *cohomology groups* $H^m(\mathfrak{A};\ \mathbb{R})$ *are isomorphic to de Rham cohomology groups* $H^m\mathscr{X}$ *of the manifold* $\mathscr{X}$.

This theorem is called the **de Rham theorem in the Leray form** (or briefly the **de Rham-Leray theorem**).

It is not hard either to point out an explicit formula giving an isomorphism

$$H^m(\mathfrak{A};\ \mathbb{R}) \to H^m\mathscr{X}. \tag{2}$$

Problem 1. If an operator $d \circ D$, where D is an operator $C^{p,q}(\mathfrak{A}) \to C^{p-1,q}(\mathfrak{A})$ defined by formula (1), is applied m times to a cochain $ic \in C^{m,0}(\mathfrak{A})$, where $c \in C^m(\mathfrak{A};\ \mathbb{R})$, then a cochain of $C^{0,m}(\mathfrak{A})$ is obtained. Show that:

1° when c is a cocycle, the cochain $(d \circ D)^m c$ is of the form $j\omega$, where ω is a closed form of $\Omega^m\mathscr{X}$;

2° the cohomology class $[\omega]$ of ω depends only on the cohomology class $[c]$ of the cocycle c;

3° up to a sign the mapping $[c] \mapsto [\omega]$ is an isomorphism (2).

Corollary 1. *For any numerable Leray covering* $\mathfrak{A}$ *of a smooth manifold* $\mathscr{X}$ *cohomology groups* $H^m(\mathfrak{A};\ \mathbb{R})$ *are* *independent, up to isomorphism, of that covering and are defined exclusively by the manifold* $\mathscr{X}$ *itself.* □

Remark 1. It can be shown, on the basis of other considerations not related to differential forms, that Corollary 1 is valid for cohomology groups $H^m(\mathfrak{A}; G)$ also over an arbitrary group G.

Definition 3. Groups $H^m(\mathfrak{A}; G)$, where $\mathfrak{A}$ is an arbitrary numerable Leray covering of a manifold $\mathscr{X}$, are called *Čech cohomology groups* of the manifold $\mathscr{X}$ and designated $H^m(\mathscr{X}; G)$. They are defined if the manifold $\mathscr{X}$ has at least one numerable Leray covering.

By virtue of Remark 1 this definition is correct. It allows Theorem 1 to be restated as follows:

Theorem 1a. *If for a manifold* $\mathscr{X}$ *there is a numerable Leray covering, then its de Rham cohomology groups are isomorphic to its Čech cohomology groups:*

$$H^m\mathscr{X} \simeq H^m(\mathscr{X}; \mathbb{R}), \quad m \geqslant 0.$$

While in practice the question of the existence of a numerable Leray covering usually causes no trouble for a given particular manifold $\mathscr{X}$, it is nevertheless interesting to consider it in the general form.

Theorem 2. *Any Hausdorff manifold satisfying the second axiom of countability (in particular, any Hausdorff compact manifold) is paracompact, i.e. each of its open coverings is numerable.*

Remark 2. It can be shown that, conversely, *any connected paracompact Hausdorff manifold satisfies the second countability axiom.*

See Remarks 2 and 3 of Lecture 24 below.

Theorem 3. *Any paracompact Hausdorff manifold has a Leray covering. If a manifold is compact, then it has a finite Leray covering.*

Corollary 1. *Čech cohomology groups* $H^m(\mathscr{X}; G)$ *are defined for any paracompact Hausdorff manifold* $\mathscr{X}$. *For a compact manifold these groups have over* G *a finite number of generators (for* $G = \mathbb{R}$ *they are finite-dimensional).*

Corollary 2. *De Rham cohomology groups* $H^m\mathscr{X}$ *of an arbitrary paracompact Hausdorff manifold* $\mathscr{X}$ *are isomorphic to its Čech groups over the field* $\mathbb{R}$:

$$H^m\mathscr{X} \simeq H^m(\mathscr{X}; \mathbb{R}).$$

Corollary 2 is generally called the **de Rham theorem for Čech cohomology groups.**

[In general, a variety of theorems describing groups $H^m\mathscr{X}$ in topological terms are known as de Rham theorems. The name is justified by the fact that the very first of these theorems has indeed been proved by de Rham. It will be discussed in Lecture 29.]

Corollary 3. *For a compact Hausdorff manifold* $\mathscr{X}$ *all vector spaces* $H^m\mathscr{X}$ *are finite-dimensional (and hence Betti numbers* $h^m\mathscr{X}$ *are defined).* □

We prove Theorem 2 in Lecture 24. As far as Theorem 3 is concerned, just as with the Gauss lemma of Lecture 5 we are compelled to put off its proof until the next semester.

Generally speaking, even for the most primitive manifolds $\mathscr{X}$ Leray coverings $\mathfrak{A}$ contain quite a lot of elements, which makes the problem of computing groups $H^m(\mathfrak{A};\ \mathbb{R}) = H^m(\mathscr{X};\ \mathbb{R})$ if not difficult, but awkward and tedious. Of interest therefore are ways of computing groups $H^m(\mathscr{X};\ \mathbb{R})$ with the aid of coverings that are not of the Leray type. (As we see from the examples of Lecture 20, this approach may turn out to be quite effective.) In view of Theorem 2 (true, not proved by us as yet) we may, being interested practically only in manifolds satisfying the second countability axiom restrict ourselves to numerable coverings $\mathfrak{A} = \{U_k,\ k \in K\}$ for which the augmented complex $C^{\cdot,\cdot}(\mathfrak{A})$ is known to be acyclic in rows, and hence its cohomology groups $H^m(\mathfrak{A})$ are isomorphic to de Rham cohomology groups $H^m\mathscr{X}$ of a manifold $\mathscr{X}$. The problem of computing groups $H^m\mathscr{X}$ is thus reduced to computing groups $H^m(\mathfrak{A})$.

We assume that

(a) we know the nerve K of a covering $\mathfrak{A}$ (i.e. we know for which indices $k_0, \ldots, k_p \in K$ the intersection $U_{k_0} \cap \ldots \cap U_{k_p}$ is not empty);

(b) for any simplex $\{k_0, \ldots, k_p\}$ of K all de Rham cohomology groups $H^q(U_{k_0} \cap \ldots \cap U_{k_p})$ of an open submanifold $U_{k_0} \cap \ldots \cap U_{k_p}$ are known;

and seek an algebraic procedure which allows us, if not to completely compute groups $H^m(\mathfrak{A}) \simeq H^m\mathscr{X}$ on the basis of these data (which continues to be an unattainable ideal) at least obtain sufficiently ample information about these groups.

Suppose $C^{\cdot,\cdot} = \{C^{p,q};\ \delta,\ d\}$ is as yet an arbitrary double complex. Since columns $C^{p,\cdot} = \{C^{p,q};\ d\}$ of $C^{\cdot,\cdot}$ are the usual cochain complexes, their cohomology groups $H^q(C^{p,\cdot})$ are defined. By tradition it is customary to designate these groups as $E_1^{p,q}$. (Note the order of the superscripts!)

Since by definition horizontal coboundary operators $\delta\colon C^{p,q} \to C^{p+1,q}$ make up a cochain mapping $C^{p,\cdot} \to C^{p+1,\cdot}$ they induce homomorphisms

$$\delta^*\colon E_1^{p,q} \to E_1^{p+1,q}. \tag{3}$$

Since $\delta \circ \delta = 0$, we have $\delta^* \circ \delta^* = 0$, and hence for every $q \geqslant 0$ the family $E_1^{\cdot,q} = \{E_1^{p,q};\ \delta^*\}$ of groups and homomorphisms is a cochain complex. (In what follows we shall write simply δ instead of δ^*.) Cohomology groups $H^p(E_1^{\cdot,q})$ of complexes $E_1^{\cdot,q}$ are, also by tradition, designated $E_2^{p,q}$ and arranged in an array of the form

(4)

$E_2^{0,q}$	$E_2^{1,q}$	$\dots$	$E_2^{p,q}$	$\dots$
. . . .	. . . .	. . . .	. . . .	. . . .
$E_2^{0,1}$	$E_2^{1,1}$	$\dots$	$E_2^{p,1}$	$\dots$
$E_2^{0,0}$	$E_2^{1,0}$	$\dots$	$E_2^{p,0}$	$\dots$

By convention $E_2^{p,q} = H_\delta^p H_d^q(C^{\cdot,\cdot})$ (and $E_2^{p,q} = H_\delta^p H_d^q(\mathfrak{A})$ if $C^{\cdot,\cdot} = C^{\cdot,\cdot}(\mathfrak{A})$), where subscripts indicate the coboundary with respect to which cohomology groups are computed.

Notice that in the special case of a double complex of the form $C^{\cdot,\cdot}(\mathfrak{A})$ *the information contained in* (a) *and* (b) (see above) *is in principle enough for groups* $E_2^{p,q}$ *to be computed.* Indeed, as we know, each column $C^{p,\cdot}(\mathfrak{A})$ of a complex $C^{\cdot,\cdot}(\mathfrak{A})$ is a direct product of de Rham complexes of the form $\Omega^{\bullet}(U_{k_0}\cap\ldots\cap U_{k_p})$ and hence its cohomology groups $E_1^{p,q}=H^q(C^{p,\bullet}(\mathfrak{A}))$ are direct products of de Rham cohomology groups $H^q(U_{k_0}\cap\ldots\cap U_{k_p})$. This means that going from group $C^{p,q}(\mathfrak{A})$ to group $E_1^{p,q}$ consists in that, restricting ourselves to cochains

$$c\colon (k_0, \ldots, k_p)\mapsto c(k_0, \ldots, k_p)\in\Omega^q(U_{k_0}\cap\ldots\cap U_{k_p})$$

whose values $c(k_0, \ldots, k_p)$ are closed forms, we go to their cohomology classes

$$[c(k_0, \ldots, k_p)]\in H^q(U_{k_0}\cap\ldots\cap U_{k_p}).$$

Consequently, we may regard as elements of $E_1^{\cdot,q}$ the skew-symmetric functions

$$\bar{c}\colon (k_0, \ldots, k_p)\mapsto\bar{c}(k_0, \ldots, k_p),$$

defined on K_p, for which $\bar{c}(k_0, \ldots, k_p)\in H^q(U_{k_0}\cap\ldots\cap U_{k_p})$ (*cochains with values in cohomologies*). Operator (3) is given as follows:

$$(\delta^*\bar{c})(k_0, \ldots, k_{p+1})=\sum_{i=0}^{p+1}(-1)^i\bar{c}(k_0, \ldots, \hat{k}_i, \ldots, k_{p+1}),$$

where omitted on the right are implicit homomorphisms of a restriction

$$H^q(U_{k_0}\cap\ldots\cap\hat{U}_{k_i}\cap\ldots\cap U_{k_p})\\ \to H^q(U_{k_0}\cap\ldots\cap U_{k_p})$$

(cf. formula (2) of the preceding lecture). All this leads to an algorithm, quite efficient in not very cumbersome cases, for computing groups $E_2^{p,q}$ □.

Thus, for double complexes of the form $C^{\bullet,\bullet}(\mathfrak{A})$ groups $E_2^{p,q}$ may be assumed to be known, at least theoretically.

Example 1. If $\mathfrak{A}=\{U, V\}$ is the two-element covering of $\mathbb{S}^n$ considered in Lecture 20, then among the groups $H^m(U)$, $H^m(V)$, and $H^m(U\cap V)$ only the groups

$$H^0(U)\simeq\mathbb{R},\quad H^0(V)\simeq\mathbb{R},\quad H^0(U\cap V)\simeq\mathbb{R}$$

and

$$H^{n-1}(U \cap V) \simeq H^{n-1}S^{n-1} \simeq \mathbb{R}$$

are nonzero. The complex $E_1^{\cdot,0}$ is therefore isomorphic to the complex $C^{\cdot}(\mathfrak{A}; \mathbb{R})$, and hence *groups* $E_2^{p,0}$ *are isomorphic to groups* $H^p(\mathfrak{A}; \mathbb{R})$. (This is a general fact valid for any manifold $\mathcal{X}$ and any one of its coverings $\mathfrak{A}$ with *connected* intersections $U_{k_0} \cap \ldots \cap U_{k_p}$.)

On the other hand, it can easily be seen that $H^0(\mathfrak{A}; \mathbb{R}) \simeq \mathbb{R}$ (again this is a general fact valid for any connected—in the obvious sense—simplicial scheme) and that $H^p(\mathfrak{A}; \mathbb{R}) = 0$ for $p > 0$ (the nerve of a covering $\mathfrak{A}$, or, more exactly, its geometric realization is a segment

(5)

and hence for $p > 1$ even the group $C^p(\mathfrak{A}; \mathbb{R})$ is zero, and for $p = 1$ any element of the group $C^1(\mathfrak{A}; \mathbb{R}) = Z^1(\mathfrak{A}; \mathbb{R})$ is uniquely defined by a number $\lambda = c(0, 1)$ and hence equal to δe, where e is a zero-dimensional cochain defined as follows: $e(0) = 0$, $e(1) = \lambda$.

Thus

$$E_2^{p,0} = \begin{cases} \mathbb{R} \text{ if } p = 0, \\ 0 \text{ if } p > 0. \end{cases}$$

As far as groups $E_1^{p,q}$ for $q > 0$ are concerned, only the group $E_1^{1,n-1} = C^1(\mathfrak{A}; \mathbb{R})$ isomorphic to $\mathbb{R}$ is nonzero. Therefore among groups $E_2^{p,q}$, $q > 0$, the only nonzero group is $E_2^{1,n-1} \simeq \mathbb{R}$.

Thus the array of groups $E_2^{p,q}$ for a covering $\{U, V\}$ of the sphere S^n is of the form

$n-1$		$\mathbb{R}$	...		...
	. . . . . .	. . . . . .	. . . . . .	. . . . . .	. . . . . .
0	$\mathbb{R}$		...		...
	0	1			

(there are zeros in all the other cells, except for the indicated two). □

It remains to learn how to go from groups $E_2^{p,q}$ to groups $H^m(C^{\cdot,\cdot})$. To do this we first describe groups $E_2^{p,q}$ directly in groups $C^{p,q}$. In order to avoid the numerous reservations, we assume that groups $C^{p,q}$ are defined also for *negative* p, q and in this case are zero:

$$C^{p,q} = 0 \quad \text{if} \quad p < 0 \quad \text{or} \quad q < 0.$$

By definition each element x of a group $E_2^{p,q}$ is a δ-cohomology class of some δ-cocycle $\bar{c} \in E_1^{p,q}$ which in turn is a d-cohomology class of some d-cocycle $c \in C^{p,q}$. (In a complex of a covering $\mathfrak{A}$ the element $\bar{c}$ is a cocycle with values in cohomologies and c is obtained from $\bar{c}$ by choosing some representative in each cohomology class $\bar{c}\,(k_0, \ldots, k_p)$.) Thus every element $x \in E_2^{p,q}$ is given by some element $c \in C^{p,q}$. We shall write $x = [c]_2$ (and $\bar{c} = [c]_1$).

For an element $[c]_2$ to be defined it is of course necessary that $dc = 0$. Besides, the element $\delta c \in C^{p+1,q}$ must be of the form dc_1, where $c_1 \in C^{p+1,q-1}$ (for the element $[c]_1 \in E_1^{p,q}$ to be a δ-cocycle). It will be more convenient for us, however, to replace d by d' and represent δc as $-d'c_1$ (which is of no fundamental significance, of course). Conversely, if $dc = 0$, then an element $[c]_1$ is defined, and if $\delta c = -d'c_1$, then $\delta\,[c]_1 = 0$ and hence an element $[c]_2$ is defined. Thus an *element* $[c]_2$ *is defined if and only if there is a two-dimensional chain* $(c,\ c_1)$ *of the form*

$$\begin{tikzcd} 0 & \\ c \arrow[u, "d'"] \arrow[r, "\delta"] & {} \\ & c_1 \arrow[u, "d_1"'] \end{tikzcd} \qquad \begin{aligned} d'c &= 0,\\ \delta c + d'c_1 &= 0. \end{aligned} \tag{6}$$

Similarly it can be shown that $[c]_2 = [c']_2$ *if and only if there are cochains* $a \in C^{p-1,q}$ *and* $b \in C^{p,q-1}$ *such that*

$$d'a = 0 \quad \textit{and} \quad c' - c = \delta a - d'b;$$

or, diagrammatically:

$$\begin{tikzcd}
0 & & \\
a \arrow[u] \arrow[r] & c' \arrow[r, dash] & c \\
 & b \arrow[u] &
\end{tikzcd} \tag{7}$$

On the other hand, we know (see formula (7) of the preceding lecture) that cocycles of a complex $C^{\bullet,\bullet}$ (with respect to the operator $\mathfrak{d}$) are chains of the form

$$\begin{tikzcd}
0 & & & & & \\
c_0 \arrow[u] \arrow[r] & {} & & & & \\
 & c_1 \arrow[u] \arrow[r] & {} & & & \\
 & & c_2 \arrow[u] \arrow[r] & \ddots & & \\
 & & & c_{m-1} \arrow[u] \arrow[r] & {} & \\
 & & & & c_m \arrow[u] \arrow[r] & 0
\end{tikzcd} \tag{8}$$

$$\begin{aligned} d'c_0 &= 0, \\ \delta c_0 + d'c_1 &= 0, \\ \delta c_1 + d'c_2 &= 0, \\ &\dots\dots\dots \\ \delta c_{m-1} + d'c_m &= 0, \\ \delta c_m &= 0, \end{aligned}$$

having the greatest possible length (or terminating at the lower row of the complex $C^{\bullet,\bullet}$). If therefore cocycle (8) has the property that $c_0 = 0, \ldots, c_{p-1} = 0$, then for its component c_p an element $[c_p]_2$ of a group $E_2^{p,m-p}$ is defined.

Thus $c \mapsto [c_p]_2$ defines some mapping

$$Z^{p,m-p} \rightarrow E_2^{p,m-p}, \tag{9}$$

where $Z^{p,m-p}$ is a subgroup of a group of $\mathfrak{d}$-cocycles $Z^m(C^{\bullet,\bullet})$ of a double complex $C^{\bullet,\bullet}$ that consists of cocycles $c = (c_0, \ldots, c_m)$ for which $c_0 = 0, \ldots, c_{p-1} = 0$. It follows directly from the definition of the operator $\mathfrak{d}$ (see formula (6) of the preceding lecture) that if cocycles $c = (0, \ldots, 0, c_p, \ldots, c_m)$ and $c' = (0, \ldots, 0, c'_p, \ldots, c'_m)$ of the group $Z^{p,m-p}$ are cohomologous, then there is a chain $(a_0, a_1, \ldots, a_p, 0, \ldots, 0)$ such that

$$\begin{tikzcd}
0 \\
a_0 \arrow[u] \arrow[r] & {} \\
& a_1 \arrow[u] \arrow[r] & \ddots \\
& & a_{p-2} \arrow[u] \arrow[r] & {} \\
& & & a_{p-1} \arrow[u] \arrow[r] & c'_p - c_p \\
& & & & a_p. \arrow[u]
\end{tikzcd}$$

$$\begin{aligned} & d'a_0 = 0, \\ & \delta a_1 + d'a_1 = 0, \\ & \ldots\ldots\ldots\ldots\ldots \\ & \delta a_{p-2} + d'a_{p-1} = 0, \\ & \delta a_{p-1} + d'a_p = c'_p - c_p, \end{aligned} \tag{10}$$

This relation becomes a relation expressible by diagram (7) for $a_0 = 0, \ldots, a_{p-2} = 0$, but in the general case it is weaker than relation (7). This forces us to introduce into consideration a quotient group $\bar{E}_2^{p,m-p}$ of a group $E_2^{p,m-p}$ with respect to the subgroup of all elements of the form $[\delta a_{p-1} + d'a_p]_2$ and a composition

$$Z^{p,m-p} - \bar{E}_2^{p,m-p} \tag{11}$$

of homomorphism (9) with a factorization homomorphism $E_2^{p,m-p} \to \bar{E}_2^{p,m-p}$. Homomorphism (11), unlike homomorphism (9), makes all coboundaries contained in $Z^{p,m-p}$ vanish and hence induces some homomorphism

$$F^{p,m-p} \to \bar{E}_2^{p,m-p}, \tag{12}$$

where $F^{p,m-p}$ is a subgroup of a group $H^m(C^{\bullet,\bullet})$ and is the image of a subgroup $Z^{p,m-p}$ under a factorization homomorphism $Z^m(C^{\bullet,\bullet}) \to H^m(C^{\bullet,\bullet})$, i.e. consists of cohomology classes $[c]$ containing a cocycle c of the form $(0, \ldots, 0, c_p, \ldots, c_m)$.

It is clear that under homomorphism (12) the subgroup $F^{p+1,m-p-1}$ of the group $F^{p,m-p}$ vanishes. This homomorphism induces therefore some homomorphism

$$F^{p,m-p}/F^{p+1,m-p-1} \to \bar{E}_2^{p,m-p}. \tag{13}$$

On the other hand, the statement that a cohomology class $[c] \in F^{p,m-p}$, $c = (0, \ldots, 0, c_p, \ldots, c_{m-p})$, vanishes under homomorphism (12) implies that there is a chain $a = (a_0, \ldots, a_p, 0, \ldots, 0)$ such that

$$\begin{aligned} d'a_0 &= 0, \\ \delta a_0 + d'a_1 &= 0, \\ &\ldots\ldots\ldots \\ \delta a_{p-2} + d'a_{p-1} &= 0, \\ \delta a_{p-1} + d'a_p &= c_p, \end{aligned}$$

$$\begin{tikzcd}
0 \\
a_0 \arrow[u] \arrow[r] & {} \\
& a_1 \arrow[u] \arrow[r] & \cdot \\
& & \cdot \\
& & {} \\
& & a_{p-2} \arrow[u] \arrow[r] & {} \\
& & & a_{p-1} \arrow[u] \arrow[r] & c_p \\
& & & & a_p . \arrow[u]
\end{tikzcd}$$

But for a cocycle $c' = c - \delta a$ the component c_p is then zero, i.e. that cocycle is in the subgroup $Z^{p+1, m-p-1}$. Since $[c] = [c']$, this proves that the subgroup $F^{p+1, m-p-1}$ is the kernel of homomorphism (12) and therefore that *homomorphism* (13) *is a monomorphism.*

Subgroups $F^{p, m-p}$ make up a series of embedded subgroups

$$(14) \quad H^m(C^{\bullet,\bullet}) = F^{0,m} \supset F^{1,m-1} \\ \supset \ldots \supset F^{m,0} \supset F^{m+1,-1} = \{0\}$$

which begins with the entire group $H^m(C^{\bullet,\bullet}) = F^{0,m}$ and ends with a zero subgroup $F^{m+1,-1} = \{0\}$.

At this point it is convenient to introduce the corresponding general terminology.

Definition 4. By a *graded Abelian group* is meant an arbitrary sequence

$$H^\bullet = \{H^0, H^1, \ldots, H^m, \ldots\}$$

of Abelian groups H^m. A *filtration* of $H^\bullet$ is a family $\{F^{p,q}\}$ of Abelian groups, where $-1 \leqslant p, q < \infty$, such that:

(i) every group $F^{p,q}$ is a subgroup of a group H^{p+q};

(ii) for any $m \geqslant 0$ there is an inclusion

$$H^m = F^{0,m} \supset F^{1,m-1} \supset \ldots \supset F^{m,0} \supset F^{m+1,-1} = \{0\}.$$

A group *adjoint* to a graded group $H^{\cdot}$ with filtration is a graded group $\mathrm{Gr}\,(H^{\cdot}) = \{\mathrm{Gr}\,(H^m)\}$ for which

$$(15) \qquad \mathrm{Gr}\,(H^m) = F^{0,m}/F^{1,m-1} \oplus \ldots \oplus F^{m,0}/F^{m+1,-1}.$$

In general, groups (15) do not yet allow groups H^m to be reconstructed, but *if a group H^m is a finite-dimensional vector space (and groups $F^{p,m-p}$ are subspaces of H^m), then the group* $\mathrm{Gr}\,(H^m)$ *is isomorphic to the group H^m.* Indeed, both H^m and $\mathrm{Gr}\,(H^m)$ are then finite-dimensional vector spaces of the same dimension (equal to the sum of the dimensions of quotient spaces $F^{p,m-p}/F^{p+1,m-p-1}$, $p = 0, \ldots, m$), and we know that vector spaces of the same dimension are isomorphic. [For an isomorphism $H^m \to \mathrm{Gr}\,(H^m)$ to be constructed, it is necessary to choose in every subspace $F^{p,m-p}$ a subspace $R^{p,m-p}$ complementary to the subspace $F^{p+1,m-p-1}$ (i.e. such that $F^{p,m-p} = F^{p+1,m-p-1} \oplus R^{p,m-p}$). Then

$$H^m = R^{0,m} \oplus R^{1,m-1} \oplus \ldots \oplus R^{m,0}$$

and factorization mappings $F^{p,m-p} \to F^{p,m-p}/F^{p+1,m-p-1}$ induce isomorphisms $R^{p,m-p} \to F^{p,m-p}/F^{p+1,m-p-1}$ making up an isomorphism $H^m \to \mathrm{Gr}\,(H^m)$. (Notice that the isomorphism $H^m \to \mathrm{Gr}\,(H^m)$ is thus constructed with a considerable arbitrariness which is in principle impossible to avoid.□

Remark 3. The statement that vector spaces H^m and $\mathrm{Gr}\,(H^m)$ are isomorphic is *true without assuming their being finite-dimensional* since the existence theorem for any subspace complementary to a subspace does not require that this should be assumed.

Indeed, let $\mathcal{V}$ be an arbitrary vector space. Recall (see Definition 5 of Lecture 1.3 and Definition 4 of Lecture 1.4) that a family (set) of vectors of $\mathcal{V}$ is said to be *linearly independent* if any of its finite subfamilies (subsets) is linearly independent, and *complete* if any vector of $\mathcal{V}$ is a linear combination of vectors of some finite subfamily (subset) of it. A linearly independent and complete family of vectors is called a *basis*. It is clear that a union of any family (chain) linearly ordered with respect to inclusion of a family (chain) of linearly independent subsets of $\mathcal{V}$ is linearly independent.

Therefore, by the Zorn lemma (see Lecture 10) there are maximal elements in the set of all linearly independent subsets of $\mathcal{V}$. Since every such element is obviously a complete subset, this proves that *every vector space $\mathcal{V}$ has a basis.* By the Zorn lemma every linearly independent subset of $\mathcal{V}$ is contained in some basis. In particular, if $\mathcal{P}$ is a subspace of $\mathcal{V}$, then any basis $\{x_\alpha\}$ of $\mathcal{P}$ is contained in some basis $\{x_\alpha, y_\beta\}$ of the entire space $\mathcal{V}$. The span $\mathcal{Q}$ of complementary vectors y_β is obviously the subset of $\mathcal{V}$ complementary to $\mathcal{P}$.

Our immediate object is to compute adjoint groups (15) for the graded group $H^\cdot(C^{\cdot,\cdot}) = \{H^m(C^{\cdot,\cdot})\}$ with filtration (14). To do this we must, first, characterize the image of every homomorphism (9) and, second, describe the transition from groups $E_2^{p,q}$ to groups $\overline{E}_2^{p,q}$. We shall deal with this in the next lecture.

Lecture 23

Groups $E_r^{p,q}$ · Spectral sequences · Spectral sequence of a double complex · Spectral sequence of a covering

---

We proceed to study groups $E_2^{p,q}$ for a double complex $C^{\cdot,\cdot} = \{C^{p,q};\ \delta,\ d\}$.

As we know (see the preceding lecture)

(a$_2$) *An element* $[c]_2 \in E_2^{p,q}$, $c \in C^{p,q}$ *is defined if and only if* $d'c = 0$ *and there is a cochain* c_1 *such that* $\delta c + d'c_1 = 0$; diagrammatically:

(1)
$$\begin{tikzcd} 0 & \\ c \arrow[u] \arrow[r] & {} \\ & c_1. \arrow[u] \end{tikzcd}$$

At the same time

(b$_2$) *The equation* $[c] = [c']_2$ *holds if and only if there are cochains* a *and* b *such that* $d'a = 0$ *and* $c' - c = \delta a - d'b$; diagrammatically:

(2)
$$\begin{tikzcd} 0 & \\ a \arrow[u] \arrow[r] & c' - c \\ & b. \arrow[u] \end{tikzcd}$$

On the other hand, the element $[c]_2$ is in the image of a homomorphism $Z^{p,m-p} \to E_2^{p,m-p}$ if and only if there is chain (1) which can be extended to a chain

(3)
$$\begin{tikzcd}
0 & & & & & \\
c \arrow[u] \arrow[r] & {} & & & & \\
& c_1 \arrow[u] \arrow[r] & \ddots & & & \\
& & & {} & & \\
& & & c_{q-1} \arrow[u] \arrow[r] & {} & \\
& & & & c_q \arrow[u] \arrow[r] & 0
\end{tikzcd}$$

$$\begin{aligned}
d'c &= 0, \\
\delta c + d'c_1 &= 0, \\
\cdots & \cdots \\
\delta c_{q-1} + d'c_q &= 0, \\
\delta c_q &= 0,
\end{aligned}$$

of length $q+1$, which goes down to the bottom row of $C^{\cdot,\cdot}$.

Among all chains of the form (1) we must thus choose chains that can be extended to a chain of the form (3). It is natural to do this by extending step by step chain (1).

To extend chain (1) by one term we consider a cochain $\delta c_1 \in C^{p+2,q-1}$

$$\begin{tikzcd}
0 & \\
c \arrow[u] \arrow[r] & {} \\
& c_1 \arrow[u] \arrow[r] & \delta c_1 .
\end{tikzcd}$$

Since $d\,\delta c_1 = \delta\, dc_1 = \pm\delta\,\delta c = 0$ and $\delta\,(\delta c_1) = 0$, for a cochain δc_1 there is a diagram

$$\begin{tikzcd}
0 & \\
\delta c_1 \arrow[u] \arrow[r] & {} \\
& 0 \arrow[u]
\end{tikzcd}$$

which shows that the cochain defines some element $[\delta c_1]_2 \in E_2^{p+2,q-1}$. Since the cochain c_1 in (1) can be replaced by any cochain of the form $c_1' = c_1 + a$, where

$d'a = 0$, the cochain δc_1 is defined up to a term δa:

$$\begin{tikzcd}
0 & \\
a \arrow[u] \arrow[r] & \delta c_1' - \delta c_1 \\
 & 0. \arrow[u]
\end{tikzcd}$$

This shows (see (b_2)) that the element $[\delta c_1]_2$ is independent of the choice of cochain c_1 and is uniquely defined by cochain c. Moreover, if $[c']_2 = [c]_2$, i.e. (again see (b_2)) if

$$\begin{tikzcd}
0 & \\
a \arrow[u] \arrow[r] & c' - c \\
 & b, \arrow[u]
\end{tikzcd}$$

then

$$\begin{tikzcd}
0 & \\
c' \arrow[u] \arrow[r] & {} \\
 & c_1', \arrow[u]
\end{tikzcd}$$

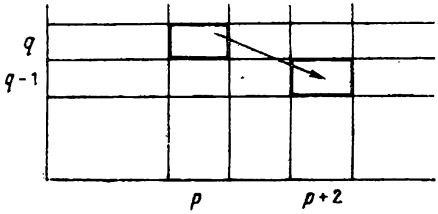
where $c_1' = c_1 + \delta b$, and therefore $\delta c_1' = \delta c_1$. Consequently, $[\delta c_1]_2$ depends only on $[c]_2$, and hence the formula

$$d_2 [c]_2 = [\delta c_1]_2$$

correctly defines some mapping

$$(4) \qquad d_2 \colon E_2^{p,q} \to E_2^{p+2,q-1};$$

diagrammatically:



It immediately follows from the definition of the mapping d_2 and statement (b_2) that

(c_2) *The equation* $d_2\,[a]_2 = [c]_2$ *is equivalent to the existence of a chain of the form*

$$\begin{tikzcd}
0 & & \\
a \arrow[u] \arrow[r] & {} & \\
& a_1 \arrow[u] \arrow[r] & c \\
& & b. \arrow[u]
\end{tikzcd} \tag{5}$$

In particular, $d_2\,[c]_2 = 0$ *if and only if there is a chain of the form*

$$\begin{tikzcd}
0 & & \\
c \arrow[u] \arrow[r] & {} & \\
& c_1 \arrow[u] \arrow[r] & {} \\
& & c_2. \arrow[u]
\end{tikzcd} \tag{6}$$

But if (5) holds, then so does (6) with $c_1 = 0$ and $c_2 = 0$. Therefore

$$d_2 \circ d_2 = 0, \tag{7}$$

i.e. for any p and q there is an inclusion

$$B_2^{p,q} \subset Z_2^{p,q},$$

where

$$Z_2^{p,q} = \operatorname{Ker}\,(d_2\colon E_2^{p,q} \to E_2^{p+2,q-1})$$

is the kernel of the mapping $d_2\colon E_2^{p,q} \to E_2^{p+2,q-1}$ and

$$B_2^{p,q} = \operatorname{Im}\,(d_2\colon E_2^{p-2,q+1} \to E_2^{p,q})$$

is the image of the mapping $d_2\colon E_2^{p-2,q+1} \to E_2^{p,q}$. We put

$$E_3^{p,q} = Z_2^{p,q}/B_2^{p,q}.$$

For an element $[c]_2 \in Z_2^{p,q}$ the corresponding element of the group $E_3^{p,q}$ (the coset $[c]_2 + B_2^{p,q}$) is designated $[c]_3$.

By definition an element $[c]_3$ is defined if and only if an element $[c]_2$ is defined and $d_2\,[c]_2 = 0$. By virtue of the second statement of (c_2) this means that

(a_3) *An element* $[c]_3 \in E_3^{p,q}$, $c \in C^{p,q}$ *is defined if and only if there is a chain of the form*

$$\begin{tikzcd}
0 & & \\
c \arrow[u] \arrow[r] & {} & \\
& c_1 \arrow[u] \arrow[r] & {} \\
& & c_2. \arrow[u]
\end{tikzcd} \tag{8}$$

According to the first statement of (c_2)

(b_3) *The equation* $[c]_3 = [c']_3$ *holds if and only if there is a chain of the form*

$$\begin{tikzcd}
0 & & & \\
a \arrow[u] \arrow[r] & {} & & \\
& a_1 \arrow[u] \arrow[r] & c' \arrow[r, dash] & c \\
& & b. \arrow[u] &
\end{tikzcd}$$

For any chain (8) the cochain δc_2 has the property that

$$\begin{tikzcd}
0 & \\
\delta c_2 \arrow[u] \arrow[r] & {} \\
& 0. \arrow[u]
\end{tikzcd}$$

An element $[\delta c_2]_2 \in E_2^{p+3,q-2}$ is therefore defined, with

$$d_2\,[\delta c_2]_2 = 0,$$

and hence so is an element $[\delta c_2]_3$. Since the difference of two chains of form (8) has the form

$$\begin{tikzcd} 0 & \\ a_1 \arrow[u] \arrow[r] & {} \\ & a_2, \arrow[u] \end{tikzcd}$$

an element $[\delta a_2]_2 = d_2\,[a_1]_2$ is added to $[\delta c_2]_2$ when chain (8) is changed, and the element $[\delta c_2]_3$ remains therefore the same. This means that the formula

$$d_3\,[c]_2 = [\delta c_2]_3$$

correctly defines some mapping

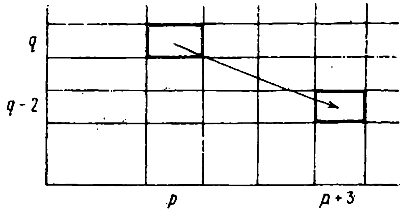
(9) $$d_3\colon Z_2^{p,q} \to E_3^{p+3,q-2}.$$

To achieve here a complete analogy with the previous case we notice that if $[c]_2 \in B_2^{p,q}$, i.e. if

$$\begin{tikzcd} 0 & & \\ a_1 \arrow[u] \arrow[r] & {} & \\ & a_2 \arrow[u] \arrow[r] & c, \end{tikzcd}$$

then for c, there is chain (8) with $c_1 = 0$ and $c_2 = 0$. Consequently, $d_3\,[c]_2 = 0$, and therefore mapping (9) induces some mapping

(10) $$d_3\colon E_3^{p,q} \to E_3^{p+3,q-2}.$$



From the definition of mapping (10) and statement (b_3) it follows immediately that (c$_3$) *The equation* $d_3\,[a]_3 = [c]_3$ *is equivalent to the existence of a chain of the form*

$$\begin{tikzcd}
0 & & & \\
a \arrow[u] \arrow[r] & {} & & \\
& a_1 \arrow[u] \arrow[r] & {} & \\
& & a_2 \arrow[u] \arrow[r] & c \\
& & & b. \arrow[u]
\end{tikzcd} \tag{11}$$

In particular, $d_3\,[c]_3 = 0$ *if and only if there are chains of the form*

$$\begin{tikzcd}
0 & & & \\
c \arrow[u] \arrow[r] & {} & & \\
& c_1 \arrow[u] \arrow[r] & {} & \\
& & c_2 \arrow[u] \arrow[r] & {} \\
& & & c_3. \arrow[u]
\end{tikzcd} \tag{12}$$

Since it follows from (11) that for a cochain c there is chain (12) with $c_1 = 0$, $c_2 = 0$, $c_3 = 0$, we see that, like the mapping d_2 the mapping d_3 satisfies the relation

$$d_3 \circ d_3 = 0,$$

and groups

$$E_4^{p,q} = Z_3^{p,q}/B_3^{p,q}$$

are therefore defined, where

$$Z_3^{p,q} = \operatorname{Ker}\,(d_3\colon E_3^{p,q} \to E_3^{p+3,q-2}),$$
$$B_3^{p,q} = \operatorname{Im}\,(d_3\colon E_3^{p-3,q+2} \to E_3^{p,q}).$$

Continuing the process, for any $r \geqslant 2$ we obtain groups $E_r^{p,q}$ consisting of elements of the form $[c]_r$, where $c \in C^{p,q}$, with the following statement holding:

(a_r) *An element* $[c]_r$ *is defined if and only if there is a chain of the form*

(13)

$$\begin{tikzcd}
0 & & & \\
c \arrow[u] \arrow[r] & {} & & \\
& c_1 \arrow[u] \arrow[r] & {} & \\
& & \ddots & \\
& & & {} \\
& & & c_{r-1}. \arrow[u]
\end{tikzcd}$$

(b_r) *The equation* $[c]_r = [c']_r$ *holds if and only if there is a chain of the form*

(14)

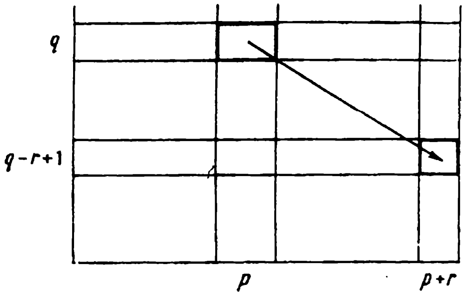
$$\begin{tikzcd}
0 & & & & \\
a_1 \arrow[u] \arrow[r] & {} & & & \\
& \ddots & & & \\
& & {} & & \\
& & a_{r-2} \arrow[u] \arrow[r] & c' \arrow[r, dash] & {} \arrow[r, dash] & c \\
& & & & b. \arrow[u] &
\end{tikzcd}$$

For any chain (13) the element $[\delta c_{r-1}]_r$ depends only on $[c]_r$, so that the formula

$$d_r\,[c]_r = [\delta c_{r-1}]_r$$

correctly defines some mapping

$$d_r\colon E_r^{p,q} \to E_r^{p+r,q-r+1}, \tag{15}$$



at the same time

(c$_r$) *the equation* $d_r\,[a]_r = [c]_r$ *is equivalent to the existence of a chain of the form*

$$\begin{tikzcd}
0 & & & \\
a \arrow[u] \arrow[r] & {} & & \\
& a_1 \arrow[u] \arrow[r] & \ddots & \\
& & a_{r-} \arrow[u, "1"'] \arrow[r] & c \\
& & & b. \arrow[u]
\end{tikzcd}$$

In particular, $d_r\,[c]_r = 0$ *if and only if there is a chain of the form*

$$\begin{tikzcd}
0 & & & \\
c \arrow[u] \arrow[r] & {} & & \\
& c_1 \arrow[u] \arrow[r] & \ddots & \\
& & c_{r-} \arrow[u, "1"'] \arrow[r] & {} \\
& & & c_r. \arrow[u]
\end{tikzcd}$$

Therefore $d_r \circ d_r = 0$, and hence groups

$$E_{r+1}^{p,q} = Z_r^{p,q}/B_r^{p,q}$$

are defined, where

$$\begin{aligned}
Z_r^{p,q} &= \operatorname{Ker}\,(d_r\colon E_r^{p,q} \to E_r^{p+r,q-r+1}),\\
B_r^{p,q} &= \operatorname{Im}\,(d_r\colon E_r^{p+r,q-r-1} \to E_r^{p,q}),
\end{aligned}$$

etc.

The constructed object deserves a special definition.

Definition 1. *By a spectral sequence* (or, more exactly, *a cohomologous spectral sequence of the first quarter*) is meant a family

$$\{E_r^{p,q};\ d_r\} \tag{16}$$

of groups (or vector spaces) $E_r^{p,q}$ and homomorphisms

$$d_r\colon E_r^{p,q} \to E_r^{p+r,q-r+1}, \tag{17}$$

where $r \geqslant 2$ and $0 \leqslant p,\ q < \infty$ (in order to avoid the numerous reservations it is convenient to assume, however, that groups $E_r^{p,q}$ are defined for negative p and q as well, with $E_r^{p,q} = 0$ if $p < 0$ or $q < 0$) that have the property that for any $r \geqslant 2$

$$d_r \circ d_r = 0 \tag{18}$$

and

$$E_{r+1}^{p,q} = Z_r^{p,q}/B_r^{p,q}, \tag{19}$$

where

$$Z_r^{p,q} = \operatorname{Ker}\,(d_r\colon E_r^{p,q} \to E_r^{p+r,q-r+1}),$$
$$B_r^{p,q} = \operatorname{Im}\,(d_r\colon E_r^{p-r,q+r-1} \to E_r^{p,q})$$

(according to (18) we have an inclusion $B_r^{p,q} \subset Z_r^{p,q}$, and hence quotient group (19) is defined).

It is customary to call homomorphisms (17) the *differentials* of a spectral sequence (16).

For every fixed $r \geqslant 2$, it is convenient to arrange groups $E_r^{p,q}$ in an array of the form

E_r:

$E_r^{p,q}$	. . .	$E_r^{p,q}$	. . .
. . . .			
$E_r^{0,0}$	. . .	$E_r^{p,0}$	. . .

This array is called the rth *term* of a spectral sequence (16).

The passage from array E_r to array E_{r+1} consists of two steps. At the first step ("scavenging") we retain in each cell $(p,\ q)$ only elements that vanish for differential (17). This step leaves the contents of the cell $(p,\ q)$ unchanged when the cell $(p + r,\ q - r + 1)$ differential (17) hits contains only zero. Since the cell $(p + r,\ q -$

$r+1$) is $r-1$ rows lower than the cell (p, q), scavenging trivially stops in (p, q), so that on the differential d_{q+1} we obtain

$$Z^{p,q}_{q+2} = Z^{p,q}_{q+3} = \dots .$$

The second step ("cancellation") consists in factoring scavenged contents of the cell (p, q) by the subgroup of elements coming from the cell $(p-r, q+r-1)$ which is r columns to the left. Hence, this step leaves the contents of the cell (p, q) unchanged when $p-r<0$, i. e. for $r \geqslant p+1$

$$B^{p,q}_{p+1} = B^{p,q}_{p+2} = \dots .$$

Thus, for any p and q the initial contents $E^{p,q}_2$ of the cell (p, q), gradually decreasing as r increases, become sooner or later *stabilized*, i.e. there is $r_0 \geqslant 2$ (namely, $r_0 = \max(p+1, q+2)$) such that

$$E^{p,q}_{r_0} = E^{p,q}_{r_0+1} = \dots .$$

We put

$$E^{p,q}_{\infty} = E^{p,q}_{r_0}.$$

By (19) each element y of a group $E^{p,q}_r$ is the image of some element of a group $E^{p,q}_{r-1}$ (that belongs to a subgroup $Z^{p,q}_{r-1}$) which in turn is the image of some element of a group $E^{p,q}_{r-2}$ (which belongs to the inverse image of the subgroup $Z^{p,q}_{r-1}$ in $E^{p,q}_{r-2}$), etc. The fact that as a result of this backward movement we get to an element x of the group $E^{p,q}_2$ is written as follows: $y = [x]_r$. When $r = r_0$ we write $[x]_\infty$ instead of $[x]_{r_0}$. (Thus, for elements $[c]_r$ of the spectral sequence constructed above from a double complex, $[x]_r$ has the same meaning as $[c]_r$ does, with c being a cochain such that $x = [c]_2$.)

If for an element $x \in E^{p,q}_2$ the element $[x]_r$ is defined, then x is said to attain the age of E_r. An element x attains the age of E_{r+1} if and only if it attains the age of E_r and $d_r [x]_r = 0$.

It is convenient (particularly in oral, informal discussions) to assume every differential d_r to be a partial and many-valued mapping from $E^{p,q}_2$ into $E^{p+r,q-r+1}_2$. [The word "partial" implies that in fact d_r is only defined on some subgroup of $E^{p,q}_2$ (namely, on the subgroup of elements that attain the age of E_r), and the word "many-valued" implies that the values of d_r are in fact not in the group $E_2^{p+r,q-r+1}$ but in some quotient group of it (or, more precisely, some subgroup of that quotient group).] Accordingly, the element $d_r[x]_r$ is generally designated as $d_r x$. It is defined if and only if $d_2 x = 0, \ldots, d_{r-1}x = 0$.

If $d_r x = 0$, then the element x is said to be a *cycle of the differential* d_r. In particular, if $d_r x = 0$ for any $r \geqslant 2$, then the element x is said to be the *cycle of all differentials.* This is the case if and only if an element $[x]_\infty$ is defined (the element x attains the age of E_∞).

The cycles of all differentials form a subgroup $Z_\infty^{p,q}$ of a group $E_2^{p,q}$. A group $E_\infty^{p,q}$ is an epimorphic image of the group $Z_\infty^{p,q}$ under a mapping

$$x \longmapsto [x]_\infty, \quad x \in Z_\infty^{p,q}.$$

Thus, if $B_\infty^{p,q}$ is the kernel of that mapping, then

$$E_\infty^{p,q} = Z_\infty^{p,q}/B_\infty^{p,q}$$

for any p and q.

Now let us return to the double complex $C^{\bullet,\bullet} = \{C^{p,q};\ \delta,\ d\}$. The problems set at the end of the preceding lecture, those of characterizing the image of a homomorphism

$$Z^{p,q} \to E_2^{p,q} \tag{20}$$

and the kernel of an epimorphism

$$E_2^{p,q} \to \overline{E}_2^{p,q}, \tag{21}$$

can be trivially solved in terms of the spectral sequence $\{E_r^{p,q};\ d_r\}$ constructed from that complex. Indeed, by definition the element $x = [c]_2$ of $E_2^{p,q}$ is in the image of (20) if and only if for a cochain c there is a chain of the form (3), i.e. (see statement (c_r)) if x is the cycle of all differentials. Hence *the image of homomorphism* (20) *is a subgroup* $Z_\infty^{p,q}$.

Similarly, the kernel of epimorphism (21) consists of elements $[c]_2$ for which there is a chain

$$\begin{tikzcd}
0 & & & & \\
a_0 \arrow[u] \arrow[r] & {} & & & \\
& a_1 \arrow[u] \arrow[r] & {} & & \\
& & \cdots & & \\
& & & a_{p-1} \arrow[u] \arrow[r] & c \\
& & & & b \arrow[u]
\end{tikzcd}$$

of length $p+1$ that begins in the first column of a complex $C^{\bullet,\bullet}$. Since, as a direct comparison of the definitions shows, these are precisely the elements of a subgroup $B_\infty^{p,q}$, this proves that

$$\bar{E}_2^{p,q} = E_2^{p,q}/B_\infty^{p,q}.$$

Hence $E_\infty^{p,q} \subset \bar{E}_2^{p,q}$ and *monomorphism* (13) *of Lecture* 22 *is an isomorphism of a quotient group* $F^{p,m-p}/F^{p+1,m-p-1}$ *onto a group* $E_\infty^{p,q}$.

The group **Gr** $(H^m(C^{\bullet,\bullet}))$ adjoint to a group $H^m(C^{\bullet,\bullet})$ is thus a direct sum of groups on the mth antidiagonal of an array E_∞:

$E_\infty^{0,m}$		...			...
	$E_\infty^{1,m-1}$	...			...
. . .	. . .	. . .	. . .	. . .	. . .
		...	$E_\infty^{m-1,1}$		...
		...		$E_\infty^{m,0}$	...

$$\mathbf{Gr}\,(H\ (C^{\bullet,\bullet})) = E_\infty^{0,m} \oplus E_\infty^{1,m-1} \oplus \ldots \oplus E_\infty^{m-1,1} \oplus E_\infty^{m,0}.$$

Definition 2. Let $H^{\cdot} = \{H^m\}$ be a graded group with filtration

$$H^m = F^{0,m} \supset F^{1,m-1} \supset \ldots \supset F^{m,0} \supset F^{m+1,-1} = \{0\}.$$

We say that a spectral sequence $\{E_r^{p,q};\ d_r\}$ *converges to the group* $H^{\cdot}$ and write

$$E_r^{p,q} \Rightarrow H^m,$$

if

$$E_\infty^{p,q} = F^{p,q}/F^{p+1,q-1}$$

for any p and q.

This definition allows the study we have made to be summed up in the following final theorem.

Theorem 1. *For any double complex* $C^{\cdot,\cdot} = \{C^{p,q};\ \delta, d\}$ *there is a spectral sequence* $\{E_r^{p,q};\ d_r\}$ *such that*

$$E_2^{p,q} = H_\delta^p\, H_d^q\, (C^{\cdot,\cdot})$$

and

$$E_r^{p,q} \Rightarrow H^m(C^{\cdot,\cdot}). \ \square$$

This theorem was in fact known to Leray (although he does not appear to have formulated it explicitly).

Corollary 1. *For any numerable covering* $\mathfrak{A}$ *of a Hausdorff smooth manifold* $\mathscr{X}$ *there is a spectral sequence*

$$\{E_r^{p,q};\ d_r\} \tag{22}$$

such that

$$E_2^{p,q} = H_\delta^p H_\delta^q(\mathfrak{A}) \tag{23}$$

and

$$E_r^{p,q} \Rightarrow H^m\mathscr{X}. \ \square \tag{24}$$

As already noted, for groups (23) to be computed it is enough to have the information contained in (a) and (b) on p. 379, and hence these groups may be assumed to be known. It is important that *as a rule we can get no other information about the spectral sequence* (22). In particular, in the general case we cannot say anything about how differentials d_r act (except that they hit a cell $(p + r, q - r + 1)$ from (p, q)) or about what filtering subgroups

$F^{p,q}$ are like. Briefly speaking, the only thing we know and can use is the naked fact that there exists a spectral sequence (22) with initial term (23) that has property (24).

It is surprising that this information should be enough in many interesting and important situations! (Though, if one stops thinking the fact over, the surprise will be gone just because the situations where this information is not sufficient are so hopeless that they lose the status of being important and interesting.)

Example 1 (continued from Example 1 of Lecture 22). As was shown, for a spectral sequence of a two-element covering $\{U, V\}$ of a sphere S^n the term E_2 is of the form

	0	1	
$n-1$		$\mathbb{R}$	...
	. . .	. . .	. . .
0	$\mathbb{R}$		...

All differentials d_r act therefore either from or into a cell containing only a zero. So they are all zero, and hence $E_2 = E_\infty$. (A spectral sequence having this property is called *degenerate*.) Thus, all antidiagonals of an array E_∞ consist of only zeros, except for the zeroth and nth ones which contain a group $\mathbb{R}$ each. Therefore

$$H^m S^n = \begin{cases} \mathbb{R} \text{ if } m=0 \text{ or } m=n, \\ 0 \text{ otherwise.} \end{cases}$$

This is the same computation as in Lecture 20, of course, but presented with the depth we have now reached.

Example 2. If $\mathfrak{U}$ is a Leray covering, then $E_1^{p,q} = 0$ and, all the more, $E_2^{p,q} = 0$ for $q > 0$. Therefore, the spectral sequence of the covering again becomes degenerate and

$$H^m \mathscr{X} = E_2^{m,0}.$$

On the other hand, it is clear that in the case under consideration the complex $\{E^{\bullet,0};\ \delta^*\}$ is nothing but a cochain complex $C^{\bullet}(\mathfrak{A};\ \mathbb{R})$ of the covering $\mathfrak{A}$, and hence

$$E_2^{m,0} = H^m(\mathfrak{A};\ \mathbb{R}) \quad \text{for any} \quad m \geqslant 0.$$

Thus we again obtain Theorem 1 of Lecture 22, which turns out to be a trivial special case of Corollary 1.

Lecture 24

Compactly exhaustible and paracompact topological spaces · Paracompact manifolds · Integrals in $\mathbb{R}^n$ · Cubable sets and densities in arbitrary manifolds · Integration of densities

---

Analytic methods of computing cohomology groups require that integral calculus should first be constructed on manifolds. We begin the construction by presenting the necessary facts from topology.

Recall (see Definition 5 of Lecture 14) that a topological space $\mathscr{X}$ is said to be *locally compact* if any point of it has a neighbourhood U whose closure $\bar{U}$ is compact.

Any Hausdorff manifold $\mathscr{X}$ *is* clearly *locally compact*. (For non-Hausdorff manifolds this is no longer the case, for example, a non-Hausdorff straight line with a singular point of infinite multiplicity is not a locally compact space.)

It is also easy to see that *if a space* $\mathscr{X}$ *satisfies the second countability axiom, is locally compact, and Hausdorff, then it has a countable base* $\{U_i\}$ *of a closure* $\bar{U}_i$ *all of whose elements are compact.* (To prove this it suffices to choose sets with compact closures in an arbitrary countable base in $\mathscr{X}$. Question for control: Why do we require that the space should be Hausdorff?)

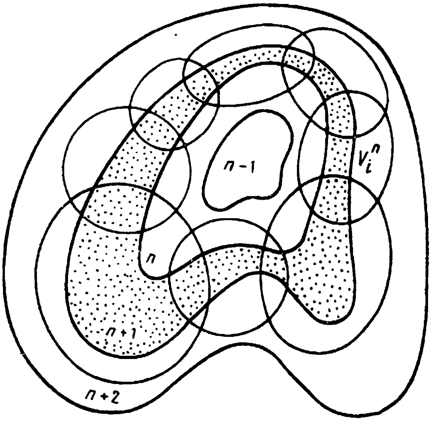
Definition 1. A topological space $\mathscr{X}$ is said to be *exhaustively compact* if it is a union of an increasing sequence

$$O_1 \subset O_2 \subset \ldots \subset O_n \subset O_{n+1} \subset \ldots$$

of open sets (called an *exhaustive compactum* of $\mathscr{X}$) such that the closure $\bar{O}_n$ of every O_n is compact and is contained in O_{n+1}.

Proposition 1. *Every Hausdorff locally compact topological space $\mathscr{X}$ satisfying the second countability axiom is exhaustively compact.*

Proof. Let $\{U_i, 1 \leqslant i < \infty\}$ be a countable base of $\mathscr{X}$ consisting of sets with compact closures. Setting $O_1 = U_1$ assume that for some $n \geqslant 2$ the set O_{n-1} has already been constructed. Since sets U_i cover $\mathscr{X}$ (and $\bar{O}_{n-1}$ is compact),



there is $j \geqslant 1$ (which for definiteness may be chosen to be the smallest) such that $\bar{O}_{n-1} \subset U_1 \cup \ldots \cup U_j$. We put $O_n = U_1 \cup \ldots \cup U_j$. Thus the sets O_n will by induction be constructed for all $n \geqslant 1$, and it is clear that they make up an exhaustive compactum of $\mathscr{X}$. □

In particular, we see that *any Hausdorff manifold satisfying the second countability axiom is exhaustively compact.*

Definition 2. An open covering $\{U_\alpha\}$ of a topological space $\mathscr{X}$ is said to be *locally finite* if any point $p \in \mathscr{X}$ has a neighbourhood that intersects with only a finite number of elements of that covering.

Recall (see Definition 1 of Lecture 8) that a covering $\{U_\alpha\}$ of $\mathscr{X}$ is a *refinement* of a covering $\{V_\beta\}$, if for any α there is β such that $U_\alpha \subset V_\beta$.

Proposition 2. *If a Hausdorff space $\mathscr{X}$ is exhaustively compact, then for any open covering $\{V_\beta\}$ of it there is a locally finite covering which is a refinement of V.*

Proof. Let $\{O_n\}$ be an arbitrary exhaustive compactum of $\mathscr{X}$. For every $n \geqslant 1$ sets $V_\beta^n = (O_{n+2} \diagdown \bar{O}_{n-1}) \cap V_\beta$ (we assume by convention that $O_0 = \varnothing$) make up an open covering of the set $\bar{O}_{n+1} \diagdown O_n$. Since $\bar{O}_{n+1} \diagdown O_n$ is compact (why?), it is possible to choose in that covering a finite subcovering $\{V_{\beta_1}^n, \ldots, V_{\beta_{m(n)}}^n\}$. Since the sets $\bar{O}_{n+1} \diagdown O_n$ cover $\mathscr{X}$, all sets of the form $V_{\beta_i}^n$, $1 \leqslant n < \infty$, $1 \leqslant i \leqslant m(n)$, make up a covering of $\mathscr{X}$ which is obviously a refinement of a covering $\{V_\beta\}$ and since every set O_k fails to intersect with sets $V_{\beta_i}^n$, $n > k$, (and hence intersects with only a finite number of sets of the form $V_{\beta_i}^n$, $n \leqslant k$), the covering $\{V_{\beta_i}^n\}$ is locally finite (for any point $p \in \mathscr{X}$ there is k such that $p \in O_k$, and hence the set O_k is a neighbourhood of that point that intersects with only a finite number of elements of $\{V_{\beta_i}^n\}$). ☐

Remark 1. In general topology a topological space $\mathscr{X}$ is said to be *paracompact* if a locally finite covering can be a refinement of any open covering of $\mathscr{X}$. In these terms Proposition 2 implies that *any Hausdorff exhaustively compact space is paracompact.*

For the case where $\mathscr{X}$ is a smooth manifold Proposition 2 can be made more precise.

An open covering $\{U_\alpha\}$ of a smooth manifold $\mathscr{X}$ is said to be *numerably compact* if there is a subordinate partition of unity $\{\eta_\alpha\}$ such that $\eta_\alpha = 0$ outside some compact set $C_\alpha \subset \mathscr{X}$.

Proposition 3. *A locally finite numerably compact covering $\{U_\alpha\}$ can be made to be a refinement of any open covering $\{V_\beta\}$ of a smooth Hausdorff manifold $\mathscr{X}$ satisfying the second countability axiom.*

Proof. Since by Proposition 1 $\mathscr{X}$ is an exhaustively compact manifold, Proposition 2 applies to it. Let $\{V_{\beta_i}^n\}$ be the covering we have constructed in proving Proposition 2. For any point $p \in V_{\beta_i}^n$ there is in $V_{\beta_i}^n$ a coordinate neighbourhood $U_{p,i,n}$ of that point with compact closure $\bar{U}_{p,i,n}$. By Proposition 2 of Lecture 14 there are open sets $V_{p,i,n}$ and $W_{p,i,n}$ such that

$$p \in W_{p,i,n}, \quad \overline{W}_{p,i,n} \subset V_{p,i,n}, \quad \overline{V}_{p,i,n} \subset U_{p,i,n},$$

and the pair $(V_{p,i,n}, W_{p,i,n})$ has a Uryson function. For every fixed i the sets of the form $W_{p,i,n}$ make up an open covering of a compact set $\overline{O}_{n+1} \setminus O_n$. On choosing in that covering a finite subcovering and doing this for all i, we denote the chosen sets $W_{p,i,n}$ by W_α. Let V_α and U_α be the corresponding sets $V_{p,i,n}$ and $U_{p,i,n}$.

By construction

(a) for any α there are inclusions

$$\overline{W}_\alpha \subset V_\alpha, \quad \overline{V}_\alpha \subset U_\alpha.$$

with the set $\overline{U}_\alpha$ (and hence each of the sets $\overline{W}_\alpha$ and $\overline{V}_\alpha$) being compact;

(b) for every pair (V_α, W_α) there is a Uryson function

$$\varphi_\alpha \colon \mathscr{X} \to \mathbb{R}$$

(recall that it has the property that $\varphi_\alpha = 1$ on $\overline{W}_\alpha$ and $\varphi_\alpha = 0$ outside V_α);

(c) the family $\{W_\alpha\}$, and hence each of the families $\{V_\alpha\}$ and $\{U_\alpha\}$, is an open covering of a space $\mathscr{X}$ (which is obviously locally finite and a refinement of a covering $\{V_\beta\}$).

Since $\{V_\alpha\}$ is a locally finite covering, the family of functions $\{\varphi_\alpha\}$ is also locally finite. Therefore a function

$$\varphi = \sum_\alpha \varphi_\alpha$$

is defined. Since $\varphi_\alpha = 1$ on $\overline{W}_\alpha$ and $\varphi_\alpha \geqslant 0$ on $\mathscr{X}$, we have $\varphi \geqslant 1$ on $\mathscr{X}$ and, in particular, $\varphi \neq 0$ on $\mathscr{X}$. Functions

$$\eta_\alpha = \frac{\varphi_\alpha}{\varphi}$$

are therefore defined. They make up a partition of unity obviously subordinate to the covering $\{U_\alpha\}$. Since $\eta_\alpha = 0$ outside the compact set $\overline{V}_\alpha$, this proves that the covering $\{U_\alpha\}$ is numerably compact. Since $\{U_\alpha\}$ is locally finite and is a refinement of the covering $\{V_\beta\}$, this completes the proof of Proposition 3. $\square$

Corollary 1. *Every open covering $\{V_\beta\}$ of a smooth Hausdorff manifold $\mathscr{X}$ satisfying the second countability axiom is numerable.*

Proof. By virtue of Proposition 3 it suffices to show that $\{V_\beta\}$ *is a numerable covering if a numerable covering* $\{U_\alpha\}$ *is a refinement of it.* Choosing for every index α an index $\beta(\alpha)$ such that $U_\alpha \subset V_{\beta(\alpha)}$ consider a partition of unity $\{\eta_\alpha\}$ subordinate to the covering $\{U_\alpha\}$. Let

$$\zeta_\beta = \sum_{\beta(\alpha)=\beta} \eta_\alpha,$$

where the summation is taken over all α for which $\beta(\alpha) = \beta$. (If there are no such α, then $\zeta_\beta = 0$.) It is clear that the function ζ_β is defined and nonnegative. Moreover, since $\zeta_\beta(p) \neq 0$ if and only if there is α such that $\beta = \beta(\alpha)$ and $\eta_\alpha(p) \neq 0$, the family of functions $\{\zeta_\beta\}$ is locally finite, and since every function η_α is a term of one and only one function ζ_β, we have

$$\sum_\beta \zeta_\beta = \sum_\alpha \eta_\alpha = 1.$$

Finally, if $p \notin V_\beta$, then $p \in U_\alpha$ for all α with $\beta(\alpha) = \beta$, and hence $\eta_\alpha(p) = 0$. Consequently, $\zeta_\beta(p) = 0$. Thus, $\{\zeta_\beta\}$ is a partition of unity subordinate to a covering $\{V_\beta\}$. Therefore this covering is a numerable covering. □

Corollary 1 means that *any smooth Hausdorff manifold satisfying the second countability axiom is paracompact.* This statement constitutes the contents of Theorem 2 of Lecture 22 which we may thus consider to be proved now.

Remark 2. It can be shown—a difficult theorem!—that *every component of a locally compact and metrizable space satisfies the second countability axiom.* On the other hand, we shall show in the next semester that *any paracompact and Hausdorff smooth manifold is metrizable. In the class of connected and Hausdorff smooth manifolds paracompactness is* therefore *equivalent to the second countability axiom.*

Remark 3. By Proposition 3 every Hausdorff smooth manifold satisfying the second countability axiom is paracompact as a topological space (see Remark 1 above). Since a topological space is paracompact if and only if so is each of its components, it follows immediately, in view of Remark 2, that a Hausdorff paracompact manifold is paracompact as a topological space as well. On the other hand, by slightly complicating the proof of Proposition 3 it can be shown that the inference of this proposition (and hence Corollary 1) remains also valid for any Hausdorff smooth manifold that is paracompact as a topological space. Consequently, a *Hausdorff manifold is paracompact if and only if it is paracompact as a topological space.*

Remark 4. The concept of a numerable covering can, of course, be immediately applied to any topological space (it suffices to replace smooth functions by continuous functions in the definition of a partition of unity), and it is easy to see that *if in a topological space $\mathscr{X}$ any open covering is numerable, then the space is paracompact* (since for every partition of unity $\{\eta_\alpha\}$ the sets $O_\alpha = \{p \in \mathscr{X};\ \eta_\alpha(p) \neq 0\}$ make up a locally finite open covering which is a refinement of any covering the partition $\{\eta_\alpha\}$ is subordinate to). It is interesting that in the class of Hausdorff spaces the converse is also true, i.e. in a *Hausdorff paracompact space every open covering is numerable.* [Notice that the statement of Remark 3 does not immediately follow from this for smooth manifolds, since numerable open coverings of a smooth manifold as a topological space a priori make up only part of its numerable open coverings.] The proof of this statement is not at all easy, and relies on a variety of difficult theorems of general topology. In the first place, we shall need the **Uryson theorem** which states that in a Hausdorff normal space there is for any pair (V, W) of open sets satisfying the relation $\overline{W} \subset V$ a continuous function equal to unity on W and to zero outside V, and Proposition 6, of Lecture 9, on the existence of contractions (which we have proved only in part). With the aid of these two theorems it can be proved quite easily that *any locally finite open covering is numerable in a Hausdorff normal space.* After that it only remains to prove the **Dieudonné theorem** (also a hard nut to crack!) according to which *every Hausdorff paracompact space is normal.*

Now we are in a position to begin constructing integral calculus on smooth manifolds.

Recall from the course of calculus that a set A of $\mathbb{R}^n$ is said to be a *null set in the Jordan sense* (or a *set of volume* 0) if for any $\varepsilon > 0$ there is a finite family of balls (or equivalently cubes or parallelepipeds) of $\mathbb{R}^n$ covering A whose total volume is smaller than ε. (Cf. the definition of null sets in the Lebesgue sense in Lecture 14 which admits countable families.)

A subset $D \subset \mathbb{R}^n$ is said to be *cubable* (or *Jordan-measurable*) if it is bounded (i.e. its closure $\bar{D}$ is compact) and its boundary (or frontier) $\operatorname{Fr} D$ is a null set. (By the *boundary* (or *frontier*) $\operatorname{Fr} D$ of a subset D of a topological space is meant a set $\bar{D} \setminus \operatorname{Int} D$.)

In order to avoid making numerous reservations we shall assume that all the functions under consideration are defined on the *entire* space $\mathbb{R}^n$. This assumption does not decrease generality, since every function f defined on a subset $D \subset \mathbb{R}^n$ may be assumed, without changing the integral, to be zero outside D, i.e. that $f(\mathbf{x}) = 0$ for $\mathbf{x} \notin D$.

A function $f: \mathbb{R}^n \to \mathbb{R}$ is said to be *finite*, if there is a compact set $C \subset \mathbb{R}^n$ such that $f = 0$ outside C, *locally bounded*, if on any compact set $C \subset \mathbb{R}^n$ this function is bounded, and *almost continuous*, if for any compact set $C \subset \mathbb{R}^n$ the function f is continuous on C outside some null set, i.e. there is a function f_C continuous on C and a null set $A \subset C$ such that $f = f_C$ on $C \setminus A$.

It is known from calculus that *any locally bounded and almost continuous function f is integrable in every cubable set D*, i.e. there is a finite integral

$$\int_D f(\mathbf{x})\, d\mathbf{x}. \tag{1}$$

Integral (1) remains unchanged if we arbitrarily change the function f on some null set or add to, or subtract from, the set D an arbitrary null set.

[We mean a Riemann integral here. If (1) is treated as Lebesgue integral, then the requirement that the set D should be cubable may be replaced by the requirement that it should be measurable. We do not need this generalization, however.]

When f is a finite function, integral (1) has the same value for every cubable set D outside which f is zero. This common value is called an *integral of f over* $\mathbb{R}^n$ and designated

$$\int_{\mathbb{R}^n} f(\mathbf{x})\, d\mathbf{x}$$

or

$$\int f(\mathbf{x})\, d\mathbf{x}, \tag{2}$$

omitting the indication of $\mathbb{R}^n$.

Since for any cubable set D and any function f integrable in D there is an equation

$$\int_D f(\mathbf{x})\,d\mathbf{x} = \int_{\mathbb{R}^n} (f\chi_D)(\mathbf{x})\,d\mathbf{x},$$

where χD is a *characteristic function* of D given as follows:

$$\chi D(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in D, \\ 0 & \text{if } \mathbf{x} \notin D, \end{cases}$$

it follows that it is possible to make use of only an integral of the form (2).

If U and V are open subsets of the space $\mathbb{R}^n$, then for any smooth mapping $\varphi\colon U \to V$ and any null set $A \subset U$ the set φA is also a null set. [The proof of this fact given in Lecture 14 for null sets in the sense of Lebesgue can be repeated word for word for null sets in the sense of Jordan.] It follows that for any cubable set $D \subset U$ with the property that $\overline{D} \subset U$, the set $\varphi D \subset V$ is also cubable. In particular, it is true when φ is a diffeomorphism. Moreover, as it is proved in calculus, for any function f integrable over φD the function $(f \circ \varphi)\,|J_\varphi|$, where

$$J_\varphi = \det \left\| \frac{\partial \varphi^i}{\partial x^j} \right\|, \quad i = j = 1, \ldots, n, \tag{3}$$

is the Jacobian of a diffeomorphism φ, is then integrable over D and there is an equation

$$\int_{\varphi D} f(\mathbf{x})\,d\mathbf{x} = \int_D (f \circ \varphi)\,|J_\varphi|\,(\mathbf{x})\,d\mathbf{x}. \tag{4}$$

This statement is known as the *theorem on the change of variables.* For integrals (2) formula (4) has the form

$$\int f(\mathbf{x})\,d\mathbf{x} = \int (f \circ \varphi)\,|J_\varphi|\,(\mathbf{x})\,d\mathbf{x}, \tag{5}$$

where f is an arbitrary finite integrable function equal to zero outside the set V, and J_φ is, by the convention agreed upon above, Jacobian (3) assumed to be zero outside U.

To construct a theory of integration on a smooth manifold $\mathscr{X}$ (which we assume to be Hausdorff and paracompact), it is first necessary to define cubable subsets in $\mathscr{X}$. This can be done without any difficulty.

Definition 3. A subset A of a smooth n-dimensional manifold $\mathscr{X}$ is said to be a *null set* if there is a finite family of charts $(U_1, h_1), \ldots, (U_N, h_N)$ such that $A \subset \bigcup U_i$

and for every $i = 1, \ldots, N$ the set $h_i(A \cap U_i)$ is a null set in $\mathbb{R}^n$ (cf. Definition 6 of Lecture 14). A subset $D \subset \mathscr{X}$ is said to be *cubable* if its closure $\overline{D}$ is compact and its boundary

$$\operatorname{Fr} D = \overline{D} \setminus \operatorname{Int} D$$

is a null set.

Problem 1. Prove that A is a null set if and only if for any finite family of charts $(U_1, h_1), \ldots, (U_N, h_N)$ that covers A (i.e. such that $A \subset \bigcup U_1$) all sets $h_i(A_i \cap U_i)$, $1 \leqslant i \leqslant N$, are null sets in $\mathbb{R}^n$.

The situation with the integrand is more delicate. For the integral to be independent of the choice of local coordinates, it is necessary that in changing the coordinates the integrand should be changed in accordance with formulas (4) or (5). This results in the following definition.

Definition 4. Suppose in each chart (U, h) of a manifold $\mathscr{X}$ some function $\rho^{(U,h)}: U \to \mathbb{R}$ (to be briefly denoted by ρ^U in what follows) is defined and for any two overlapping charts (U, h) and (V, k) there is an equation

$$\rho^V = \rho^U \left| \det \frac{\partial k}{\partial h} \right| \quad \text{on} \quad U \cap V, \tag{6}$$

where $\det \frac{\partial k}{\partial n}$ is the Jacobian of a diffeomorphism $\varphi = k \circ h^{-1}$ (interpreted as a function on $U \cap V \subset U$, i.e. related to its Jacobian J_φ in $\mathbb{R}^n$ by the formula $\det \frac{\partial k}{\partial h} = J_\varphi \circ h$). Then the family $\{\rho^U\}$ is said to be a *density* on $\mathscr{X}$.

Example 1. An arbitrary differential form ω of degree n on a manifold $\mathscr{X}$ (where as ever $n = \dim \mathscr{X}$) is, in every chart $(U, h) = (U, x^1, \ldots, x^n)$, of the form

$$\omega = w^U dx^1 \wedge \ldots \wedge dx^n,$$

where w^U is some smooth function on U. For any two charts (U, h) and (V, k) the corresponding functions w^U and w^V are connected by the formula

$$w^V = w^U \det \frac{\partial k}{\partial h} \quad \text{on} \quad U \cap V.$$

Consequently, setting $\rho^U = |w^U|$ we define on $\mathscr{X}$ some density.

A density $\rho = \{\rho^U\}$ is said to be *smooth*, if all functions ρ^U are smooth, *locally bounded*, if the functions ρ^U are locally bounded, and *almost continuous*, if all ρ^U are almost continuous.

Locally bounded and almost continuous densities on a manifold $\mathscr{X}$ form a vector space $\Pi\mathscr{X}$ which is naturally a module over a ring of all locally bounded and almost continuous functions on $\mathscr{X}$. Similarly, the vector space $\Pi_{\text{sm}}\mathscr{X}$ of all smooth densities is a module over a ring of smooth functions.

A density ρ is said to be *zero* (respectively, *positive*) at a point $p \in \mathscr{X}$ if for any coordinate neighbourhood U containing the point p there is an equation $\rho^U(p) = 0$ (respectively, an equation $\rho^U(p) > 0$). Since the function $\det \frac{\partial k}{\partial h}$ is nowhere vanishing, it follows from relation (6) that this definition is correct (is independent of the choice of coordinate neighbourhood U).

[Notice that there is no sense in speaking of the value of ρ at a point p.]

A density on a manifold $\mathscr{X}$ is said to be *finite* if there is a compact set $C \subset \mathscr{X}$ such that $\rho = 0$ outside C. Finite densities form a subspace (submodule) $\Pi^{\text{fin}}\mathscr{X}$ of a space (module) $\Pi\mathscr{X}$. On a compact (and only on a compact!) manifold $\mathscr{X}$ every density is finite.

A density positive at every point $p \in \mathscr{X}$ is called a *volume density* on $\mathscr{X}$. By tradition we use dv to denote a volume density. For every density ρ the formula

$$f = \frac{\rho^U}{dv^U} \quad \text{on} \quad U$$

(with a given volume density dv) correctly defines on $\mathscr{X}$ a function f with the property that $\rho = f\,dv$ on $\mathscr{X}$. Consequently, *if on a manifold $\mathscr{X}$ there is a smooth volume density, then vector spaces $\Pi\mathscr{X}$ and $\Pi_{\text{sm}}\mathscr{X}$ are one-dimensional free modules (over a ring of locally bounded and almost continuous functions and, respectively, over a ring of smooth functions).* [But their submodules consisting of finite densities are in general neither free nor one-dimensional.]

For a given volume density the integral

$$\int_D dv$$

is called the *volume* of a cubable set D.

Example 2. Let $\mathscr{X}$ be a two-dimensional submanifold of a Euclidean space (a surface). Then (see Lecture 3) in every coordinate neighbourhood $U \subset \mathscr{X}$ (which is an elementary surface in the sense of Definition 2 of Lecture 3) functions E, F, and G (the coefficients of the first quadratic form) are defined with the property that in transforming the coordinates the determinant

$$\begin{vmatrix} E & F \\ F & G \end{vmatrix} = EG - F^2 > 0$$

is multiplied by a square of the determinant of the transition matrix. Hence the formula

$$\rho^U = \sqrt{EG - F^2}$$

gives on $\mathscr{X}$ some volume (or rather surface) density. This is designated $d\sigma$ (or in some other similar way, for example dS) and called an *element* of the surface $\mathscr{X}$. (Cf. Lecture 3.)

Problem 2. Prove that there is a smooth volume density dv on every Hausdorff paracompact smooth manifold $\mathscr{X}$.

Thus, densities can be identified with functions on a Hausdorff paracompact manifold by choosing some volume density.

Now we are in a position to state the basic existence and uniqueness theorem for the integral. We shall even do this in two versions.

Let $\mathscr{X}$ be a smooth Hausdorff paracompact manifold.

Theorem 1. *For every cubable set $D \subset \mathscr{X}$ there is a unique linear functional*

$$\int_D : \Pi\mathscr{X} \to \mathbb{R} \tag{7}$$

with the following two properties:

(a) *If* $D = D_1 \sqcup D_2$ *(i.e. if* $D = D_1 \cup D_2$ *and* $D_1 \cap D_2 = \varnothing$*), then*

$$\int\limits_D \rho = \int\limits_{D_1} \rho + \int\limits_{D_2} \rho$$

for any density $\rho \in \Pi\mathscr{X}$*.*

(b) *If for a density* $\rho \in \Pi\mathscr{X}$ *there is a chart* (U, h) *such that* $\rho = 0$ *outside* U*, then*

$$\int\limits_D \rho = \int\limits_{h(D\cap U)} (\rho^U \circ h^{-1})(\mathbf{x})\, d\mathbf{x},$$

where on the right is an integral in $\mathbb{R}^n$*.*

Theorem 1a. *There is a unique linear functional*

$$\int : \Pi^{\text{fin}}\mathscr{X} \to \mathbb{R} \tag{8}$$

with the property that

$$\int \rho = \int\limits_{h(U)} (\rho^U \circ h^{-1})(\mathbf{x})\, d\mathbf{x} \tag{9}$$

for any chart (U, h) *and any finite density* ρ *which is zero outside* U*.*

Functional (8) is also designated as $\int\limits_{\mathscr{X}}$.

To derive Theorem 1a from Theorem 1 it suffices to put for any density $\rho \in \Pi^{\text{fin}}\mathscr{X}$

$$\int \rho = \int\limits_D \rho, \tag{10}$$

where D is an arbitrary cubable set outside which the density ρ is zero. (It is clear that such a set does exist and that integral (10) is independent of the choice of it.) Conversely, since for any density $\rho \in \Pi\mathscr{X}$ and any cubable set $D \subset \mathscr{X}$ the density $\chi_D\, \rho$, where χ_D is the characteristic function of D, is obviously a finite locally bounded and almost continuous density, putting

$$\int\limits_D \rho = \int \chi_D \rho$$

we construct integral (7) from integral (8). (Property (b) for integral (7) is trivial, and property (a) follows from the fact that $\chi_{D_1 \sqcup D_2} = \chi_{D_1} + \chi_{D_2}$.) Theorems 1 and 1a thus follow from each other, and, therefore only one of them needs to be proved. We prove Theorem 1a.

Proof of Theorem 1a. As ever, we first prove uniqueness. Since under the hypothesis the manifold $\mathscr{X}$ is Hausdorff and paracompact, it has a numerable locally finite covering $\{U_\alpha\}$ consisting of coordinate neighbourhoods. Let $\{\eta_\alpha\}$ be a partition of unity subordinate to $\{U_\alpha\}$. Then for any density $\rho \in \Pi^{\text{fin}}\mathscr{X}$ only a finite number of densities $\eta_\alpha\rho$ are nonzero (prove this!), and hence there is a formula

$$\rho = \sum_\alpha \eta_\alpha \rho.$$

Since $\eta_\alpha\rho = 0$ outside U_α, it follows (see (9)) that

$$\int \rho = \sum_\alpha \int\limits_{h_\alpha(U_\alpha)} (\eta_\alpha \rho^{U_\alpha} \circ h_\alpha^{-1})(\mathbf{x})\, d\mathbf{x}, \tag{11}$$

where h_α is a coordinate mapping $U_\alpha \to \mathbb{R}^n$. Since the right-hand side of (11) is independent of the choice of functional (8), this proves that the functional is unique.

To prove existence we define functional (8) by formula (11). It is clear that this functional is linear. Moreover, if $\rho = 0$ outside U, then by the theorem on the change of variables (see (4))

$$\begin{aligned}\int\limits_{h_\alpha(U_\alpha)} (\eta_\alpha \rho^{U_\alpha} \circ h_\alpha^{-1})(\mathbf{x})\, d\mathbf{x} &= \int\limits_{h_\alpha(U_\alpha \cap U)} (\eta_\alpha \rho^{U_\alpha} \circ h_\alpha^{-1})(\mathbf{x})\, d\mathbf{x} \\ &= \int\limits_{h(U_\alpha \cap U)} (\eta_\alpha \rho^{U_\alpha} \circ h^{-1})\, |J_\varphi|(\mathbf{x})\, d\mathbf{x},\end{aligned}$$

where $\varphi = h \circ h_\alpha^{-1}$. Since by definition

$$J_\varphi = \det \frac{\partial h}{\partial h_\alpha} \circ h^{-1}$$

and

$$(\eta_\alpha \rho^{U_\alpha} \circ h^{-1}) \left| \det \frac{\partial h}{\partial h_\alpha} \circ h^{-1} \right| = \eta_\alpha \rho^U \circ h^{-1}$$

(see (6)), it follows that

$$\int\limits_{h_\alpha(U_\alpha)} (\eta_\alpha \rho^{U_\alpha} \circ h_\alpha^{-1})(\mathbf{x})\, d\mathbf{x} = \int\limits_{h(U_\alpha \cap U)} (\eta_\alpha \rho^U \circ h^{-1})(\mathbf{x})\, d\mathbf{x}$$

$$= \int\limits_{h(U)} (\eta_\alpha \rho^U \circ h^{-1})(\mathbf{x})\, d\mathbf{x}$$

and hence

$$\int \rho = \sum_\alpha \int\limits_{h(U)} (\eta_\alpha \rho^U \circ h^{-1})(\mathbf{x})\, d\mathbf{x}$$

$$= \left(\sum_\alpha \eta_\alpha\right) \int\limits_{h(U)} (\rho^U \circ h^{-1})(\mathbf{x})\, d\mathbf{x} = \int\limits_{h(U)} (\rho^U \circ h^{-1})(\mathbf{x})\, d\mathbf{x}. \square$$

Given a volume density dv on a manifold $\mathscr{X}$, for any finite locally bounded and almost continuous function f on $\mathscr{X}$ an integral

$$\int f\, dv = \int\limits_{\mathscr{X}} f\, dv$$

called an *integral of f over $\mathscr{X}$ with respect to dv* is defined.

Example 3 (continued from Example 2). For any finite (and, say, smooth) function f on a surface $\mathscr{X}$ of a Euclidean space an integral

$$\int\limits_{\mathscr{X}} f\, d\sigma, \tag{12}$$

where $d\sigma$ is a surface element on $\mathscr{X}$ (see Example 2 above) is defined. Integrals of the form (12) are called *surface integrals of the first kind.* It is shown in calculus that they are limits of naturally defined integral sums resulting from partitioning of a surface into a large number (let it in what follows tend to infinity) of surface elements. [Cf. remarks on measuring surface areas.]

Example 4. Let γ: $[a, b] \to \mathbb{R}^n$ be an arbitrary smooth curve in a Euclidean space $\mathbb{R}^n$, which is not in general subject to any regularity requirements, i.e. in other words an arbitrary smooth vector function

$$\mathbf{r} = \mathbf{r}(t), \quad a \leqslant t \leqslant b$$

(see Lecture 1). For any function $f = f(\mathbf{r})$ on $\mathbb{R}^n$ whose domain contains the support $\gamma[a, b]$ of the curve γ we put

$$\int_{\gamma} f\,ds = \int_a^b f(\mathbf{r}(t))\,|\mathbf{r}'(t)|\,dt. \tag{13}$$

This integral is called a *curvilinear* (or *line*) *integral of the first kind* of the function f over the curve γ. It is a limit of integral sums of the form $\sum f(\mathbf{r}_i)\,s_i$, where s_i are the lengths of segments in an arbitrary partition of the curve γ, and $\mathbf{r}_i$ are some of their points. By the general definitions introduced earlier this integral is an integral over a smooth manifold $\mathscr{X} = (a, b)$ of a function $f \circ \gamma\colon t \mapsto f(\mathbf{r}(t))$ with respect to the density ds given in a chart $(\mathscr{X}, \mathrm{id})$ by the formula $t \mapsto |\mathbf{r}'(t)|$. In the special case where the curve γ (or, more precisely, its restriction to (a, b)) is simple and regular, i.e. where its support $\mathscr{L}$ is a one-dimensional embedded submanifold and the pair $(\mathscr{L}, \gamma^{-1})$ is a chart on $\mathscr{L}$, integral (13) may also be interpreted as an integral over $\mathscr{L}$ of the function f (more precisely, of its restriction $f|_{\mathscr{L}}$ to $\mathscr{L}$) with respect to the volume density ds (now called a *line element*) given in the chart $(\mathscr{L}, \gamma^{-1})$ by the function $t \mapsto |\mathbf{r}'(t)|$.

Lecture 25

Orientable manifolds · Integration of forms · Poincaré lemma for finite forms · The group $H^n_{\text{fin}}\mathscr{X}$ · An orientable manifold

---

To apply integral calculus to computing de Rham cohomology groups, we must learn how to integrate forms rather than densities. To do this it is necessary to apply the notion of orientation of a vector (or affine) space, known to us from the first semester, to arbitrary smooth manifolds.

Definition 1. Charts $(U, h) = (U, x^1, \ldots, x^n)$ and $(U', h') = (U', x^{1'}, \ldots, x^{n'})$ of a smooth n-dimensional ($n > 0$) manifold $\mathscr{X}$ is said to be *positively oriented* if either $U \cap U' = \varnothing$ or $U \cap U' \neq \varnothing$ and

$$\det \frac{\partial h'}{\partial h} > 0 \text{ on } U \cap U',$$

i.e. if at every point $p \in U \cap U'$ bases

$$\left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^n}\right)_p \text{ and } \left(\frac{\partial}{\partial x^{1'}}\right)_p, \ldots, \left(\frac{\partial}{\partial x^{n'}}\right)_p$$

of a tangent space $\mathrm{T}_p\mathscr{X}$ are of the same sign. An atlas consisting of positively oriented charts is called an *orienting atlas.* A manifold $\mathscr{X}$ with at least one orienting atlas is called *orientable.*

It is clear that *a manifold is orientable if and only if so are all its components.*

It can easily be seen (cf. Corollary 1 to Proposition 1 of Lecture 6) that for any orienting atlas A of an orientable manifold $\mathscr{X}$ the set $\mathsf{A}^+_{\max}$ of all charts positively oriented with respect to each chart of A is an orienting atlas containing A, and it is a maximal atlas (i.e. one with the property that if an orienting atlas A^* contains the atlas A, then $\mathsf{A}^* \subset \mathsf{A}^+_{\max}$).

Problem 1. Prove that orientable are

(a) spheres $\mathbb{S}^n$, $n \geqslant 0$,

(b) projective spaces $\mathbb{R}P^{2n+1}$ of odd dimensions;

(c) realizations of complex analytic manifolds (see Lecture 11);

(d) Lie groups.

[As far as projective spaces of even dimension are concerned, and the projective plane $\mathbb{R}P^2$ in particular, it can be shown (a hard nut to crack!) that these spaces are nonorientable.]

Definition 2. An orientable manifold in which a maximal orienting atlas is chosen is said to be *oriented*, and the chosen atlas is its *orientation*. Charts in the orientation of an oriented manifold are called *positive*.

For every point $p \in \mathscr{X}$ the orientation of $\mathbf{T}_p\mathscr{X}$ defined by the basis

$$\left(\frac{\partial}{\partial x^1}\right)_p, \ldots, \left(\frac{\partial}{\partial x^n}\right)_p$$

is by definition the same for all positive charts $(U, x^1, \ldots, x^n)$. This orientation is said to be *induced* by a given orientation of the manifold $\mathscr{X}$.

Thus the orientation of a manifold is obviously a choice of compatible orientations of the spaces tangent to the manifold.

Let $(U, h) = (U, x^1, x^2, \ldots, x^n)$ be an arbitrary chart in an oriented manifold $\mathscr{X}$. A point $p \in U$ is said to be *positive* if for one (and therefore for every) positive chart (V, k) with the property that $p \in V$ there is an inequality

$$\det \frac{\partial h}{\partial k} > 0.$$

It is clear that both the set of all positive points and its complement are open in U. Therefore, if (U, h) is *connected* (i.e. the set U is connected), then either all the points of U are positive, and hence (U, h) is positive, or there are no positive points in U at all and then a chart $(U, -x^1, x^2, \ldots, x^n)$ is positive. Thus, *for any con-*

nected chart $(U, x^1, x^2, \ldots, x^n)$ *of an oriented manifold* $\mathcal{X}$ *one (and only one) of the two charts* $(U, x^1, x^2, \ldots, x^n)$ *and* $(U, -x^1, x^2, \ldots, x^n)$ *is positive.*

Problem 2. Deduce from this that *there are two and only two orientations in a connected orientable manifold of dimension* $n > 0$. [These orientations are called *opposite.* If $\mathcal{X}$ is a manifold with one orientation, then given an opposite orientation it is usually denoted by $-\mathcal{X}$.]

It is clear that orientations on different components of an orientable manifold $\mathcal{X}$ may be given independently of each other. There are therefore exactly 2^N different orientations on a manifold $\mathcal{X}$ with N components.

The place orientable manifolds occupy in the theory of integration is determined by the following proposition.

Proposition 1. *There is a natural bijective correspondence between smooth densities and smooth differential forms of degree n on an oriented n-dimensional* $(n > 0)$ *manifold* $\mathcal{X}$.

Proof. Let ρ be an arbitrary smooth density on $\mathcal{X}$. For every positive chart $(U, h) = (U, x^1, \ldots, x^n)$ we define on U a form ω_U of degree n as follows:

$$\omega_U = \rho^U dx^1 \wedge \ldots \wedge dx^n.$$

Since for any two positive (and hence positively oriented) charts (U, h) and (U', h') we have

$$\rho^{U'} = \rho^U \det \frac{\partial h'}{\partial h} \text{ on } U \cap U',$$

there is $\omega_U = \omega_{U'}$, on $U \cap U'$, and hence

$$\omega = \omega_U \quad \text{on} \quad U$$

correctly defines a differential form ω of degree n on $\mathcal{X}$.

Conversely, let ω be an arbitrary differential form of degree n on $\mathcal{X}$, $(U, h) = (U, x^1, x^2, \ldots, x^n)$ be an arbitrary chart in $\mathcal{X}$, and let

$$\omega = w^U dx^1 \wedge \ldots \wedge dx^n \text{ on } U.$$

We define on U a function ρ^U, assuming that on every component U' of the set U

$$(1)\qquad \rho^U = \begin{cases} w^U & \text{if } (U', x^1, x^2, \ldots, x^n) \text{ is a positive chart,} \\ -w^U & \text{otherwise (i.e. if } (U', -x^1, x^2, \ldots, x^n) \text{ is a positive chart).} \end{cases}$$

An obvious check suggests that functions ρ^U make up a density and that the constructed correspondences $\rho \mapsto \omega$ and $\omega \mapsto \rho$ are reciprocal. □

Forms of degree n on an n-dimensional manifold are also known as *forms of maximal degree.*

For any form ω of degree n on an n-dimensional oriented (paracompact and Hausdorff) manifold $\mathscr{X}$ and any cubable set D we put

$$\int_D \omega = \int_D \rho,$$

where ρ is density (1) corresponding to a form ω. Similarly, if ω is *finite* (i.e. zero outside some compact set or, equivalently, if the density ρ corresponding to that form is finite), then by definition

$$\int \omega = \int \rho.$$

Just as for densities, both integrals reduce directly to each other: for any finite form ω

$$\int \omega = \int_D \omega,$$

where D is an arbitrary cubable set outside which the form ω is zero, and conversely

$$\int_D \omega = \int \chi_D \omega$$

for any form ω and any cubable set D.

For integrals of forms formula (11) of Lecture 24 is as follows:

$$(2)\qquad \int \omega = \sum_\alpha \int_{h_\alpha(U_\alpha)} (\eta_\alpha w^U \circ h_\alpha^{-1})(\mathbf{x})\, d\mathbf{x},$$

where $\{(U_\alpha, h_\alpha)\}$ is an arbitrary numerable locally finite covering consisting of positive charts and $\{\eta_\alpha\}$ is the subordinate partition of unity.

Note that $\int_D \omega$ and $\int \omega$ *depend* on the orientation of the manifold $\mathscr{X}$ (and on the connected manifold, for example, they change sign when passing to the opposite orientation).

Problem 3. Given on a manifold $\mathscr{X}$ a family of forms $\{\omega_t\}$ of degree n which are smoothly (or continuously) dependent on a parameter t, show that the integral

$$\int \omega_t$$

is then a smooth (or, respectively, continuous) function of t.

Remark 1. We assumed throug out the foregoing that $n > 0$. In the degenerate case of $n = 0$ the manifold $\mathscr{X}$ is a set of isolated points, and a special definition of orientation is necessary here. We shall say that a zero-dimensional manifold $\mathscr{X}$ is *oriented* if each of its points p is associated with a sign $\varepsilon(p) = \pm 1$. Forms ω of maximal degree on the zero-dimensional manifold $\mathscr{X}$ are arbitrary functions f on $\mathscr{X}$, and finite forms are functions that are nonzero at only a finite number of points $p \in \mathscr{X}$. The integral of such a function over an oriented zero-dimensional manifold $\mathscr{X}$ is defined as follows:

$$\int f = \sum_{p \in \mathscr{X}} \varepsilon(p) f(p), \tag{3}$$

which makes sense because of the assumption of finiteness.

We first apply integrals of forms to compute the group $H^n\mathscr{X}$ of an arbitrary orientable n-dimensional **compact** smooth Hausdorff manifold $\mathscr{X}$. In fact we shall compute, assuming $\mathscr{X}$ to be only paracompact, not the group $H^n\mathscr{X}$ but some other group $H^n_{\text{fin}}\mathscr{X}$ which coincides with $H^n\mathscr{X}$ when $\mathscr{X}$ is compact. To do this we must first apply the Poincaré lemma to finite forms given on $\mathbb{R}^n$.

Every differential form ω of degree m on $\mathbb{R}^n$ can be expressed as follows:

$$(4)\qquad \omega = \sum_{1 \leqslant i_1 < \ldots < i_m \leqslant n} \cdots \sum w_{i_1 \ldots i_m}\, dx^{i_1} \wedge \ldots \wedge dx^{i_m},$$

where $w_{i_1 \ldots i_m}$ are smooth functions on $\mathbb{R}^n$. We shall consider *families* $\{\omega_p\}$ of such forms which depend on a point p of some smooth manifold $\mathscr{X}$. Each form ω_p of such a family can be expressed by formula (4) with coefficients dependent on p, i.e. those that are functions on the product $\mathbb{R}^n \times \mathscr{X}$. We shall always assume that *all these functions are smooth*. [Such families $\{\omega_p\}$ can naturally be identified with forms on $\mathbb{R}^n \times \mathscr{X}$ dependent only on the differentials of coordinates $x^1, \ldots, x^n$ (cf. Lecture 20), but we shall not need this identification here.]

We shall be interested mainly in families of finite forms of maximal degree n. Every such form ω_p is as follows:

$$(5)\qquad \omega_p = w_p\, dx^1 \wedge \ldots \wedge dx^n,$$

where w_p is a smooth function of $\mathbf{x}$ (and p) equal to zero outside some cube

$$\mathring{\mathbf{I}}^n_r = \{\mathbf{x} \in \mathbb{R}^n\ ;\ |x^1| < r, \ldots, |x^n| < r\}$$

whose side $2r$ is in general dependent on p), with

$$(6)\qquad \int_{\mathbb{R}^n} \omega_p = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} w_p(\mathbf{x})\, d\mathbf{x}.$$

Proposition 2. *If integral* (6) *is identically* (*over* p) *zero for a finite form* (5), *then on* $\mathbb{R}^n$ *there are finite forms* θ_p *of degree* $n-1$, *which smoothly depend on* $p \in \mathscr{X}$, *such that*

$$d\theta_p = \omega_p \textit{ for any point } p \in \mathscr{X}.$$

Besides it may be assumed that if $\omega_p = 0$ *outside* $\mathring{\mathbf{I}}^n_r$, *then* $\theta_p = 0$ *outside* $\mathring{\mathbf{I}}^n_r$ *as well* (*in particular*, *if* $\omega_p = 0$, *then* $\theta_p = 0$).

Proof. We proceed by induction on n. When $n = 1$ the form ω_p can be expressed as follows:

$$\omega_p = w_p\, dx,$$

where $w_p = w_p(x)$ is a smooth (and smoothly p-dependent) function of $x \in \mathbb{R}$ which is zero outside some interval $(-r, r)$. Let

$$\theta_p(x) = \int\limits_{-r}^{x} w_p(x)\, dx.$$

The function θ_p is smoothly dependent on p and $d\theta_p = w_p$. Since $\int \omega_p = \theta_p(r)$, under the hypothesis $\theta_p(r) = 0$, and hence $\theta_p(x) = 0$ for $x \geqslant r$. Since obviously $\theta_p(x) = 0$ for $x \leqslant -r$, it follows that $\theta_p(x) = 0$ for $|x| \geqslant r$. This proves Proposition 1 for $n = 1$.

Assuming now that Proposition 1 has already been proved for families of forms of degree $n - 1$ on the space $\mathbb{R}^{n-1}$, consider a family of forms (5) on $\mathbb{R}^n$. Let $p \in \mathscr{X}$ and let r be a (p-dependent) number such that $\omega_p = 0$ outside $\mathring{\mathbf{I}}_r^n$.

Representing the space $\mathbb{R}^n$ as a product $\mathbb{R}^{n-1} \times \mathbb{R}$ and identifying accordingly its points with pairs of the form $(\mathbf{x}, t)$, where $\mathbf{x} \in \mathbb{R}^{n-1}$ and $t \in \mathbb{R}$, we can write every form ω_p as

$$\omega_p = \omega_{t,p} \wedge dt,$$

where

$$\omega_{t,p} = w_p(\mathbf{x}, t)\, dx^1 \wedge \ldots \wedge dx^{n-1},$$
$$\mathbf{x} = (x^1, \ldots, x^{n-1}),$$

is a form on $\mathbb{R}^{n-1}$ which is smoothly dependent on the point (t, p) of a manifold $\mathbb{R} \times \mathscr{X}$ and zero outside the cube $\mathring{\mathbf{I}}_r^{n-1}$ and for $|t| \geqslant r$.

With the aid of the function α in Lemma 1 of Lecture 1 a smooth nonnegative function α on $\mathbb{R}^{n-1}$ can immediately be constructed such that $\alpha = 0$ outside $\mathring{\mathbf{I}}_r^{n-1}$ and

$\alpha(0) \neq 0$. For such a function the integral

$$\int\limits_{\mathbb{R}^{n-1}} \alpha(\mathbf{x})\, d\mathbf{x} \tag{7}$$

is positive, and dividing the function α by the integral we obtain a function α_0 on $\mathbb{R}^{n-1}$ for which integral (7) is equal to unity. For the function $g_{t,p}$ on $\mathbb{R}^{n-1}$ defined by

$$g_{t,p}(\mathbf{x}) = w_p(\mathbf{x}, t) - \alpha_0(\mathbf{x}) \int\limits_{\mathbb{R}^{n-1}} w_p(\mathbf{x}, t)\, d\mathbf{x}$$

we will have

$$\int\limits_{\mathbb{R}^{n-1}} g_{t,p}(\mathbf{x})\, d\mathbf{x} = 0.$$

The form

$$\tilde{\omega}_{t,p} = g_{t,p}\, dx^1 \wedge \ldots \wedge dx^{n-1}$$

on $\mathbb{R}^{n-1}$ is zero outside $\mathring{\mathbf{I}}_r^{n-1}$ (and for $|t| \geqslant r$), smoothly depends on (t, p), and the integral of this form over $\mathbb{R}^{n-1}$ is identically zero. Hence, under the induction hypothesis there is on $\mathbb{R}^{n-1}$ a form $\theta_{t,p}$ smoothly dependent on $(t, p) \in \mathbb{R} \times \mathscr{X}$ and zero outside the cube $\mathring{\mathbf{I}}_r^{n-1}$ (and for $|t| \geqslant r$) such that

$$d\theta_{t,p} = \tilde{\omega}_{t,p}.$$

The forms $\theta_{t,p}$ and $\tilde{\omega}_{t,p}$ may be thought of as forms on $\mathbb{R}^n$ independent of dt. We shall denote them in this case by ϑ_p and $\tilde{\omega}_p$, respectively. The form $\tilde{\omega}_p$ is connected with the form ω_p by

$$\omega_p = \tilde{\omega}_p \wedge dt + f_p\, dx^1 \wedge \ldots \wedge dx^{n-1} \wedge dt,$$

where f_p is a function on $\mathbb{R}^n$ defined as follows:

$$f_p(\mathbf{x}, t) = \alpha_0(\mathbf{x}) \int\limits_{\mathbb{R}^{n-1}} \omega_p(\mathbf{x}, t)\, d\mathbf{x} = w_p(\mathbf{x}, t) - g_{t,p}(\mathbf{x}),$$

and the differential $d\vartheta_p$ of ϑ_p is the sum of $\tilde{\omega}_p$ and a certain form $\beta \wedge dt$. Therefore

$$d\vartheta_p \wedge dt = \tilde{\omega}_p \wedge dt$$

and hence

$$d\,(\vartheta_p \wedge dt) = \tilde{\omega}_p \wedge dt.$$

We define the form θ_p on $\mathbb{R}^n$ as:

$$\theta_p = \vartheta_p \wedge dt + (-1)^{n-1} F_p dx^1 \wedge \dots \wedge dx^{n-1},$$

where F_p is a function on $\mathbb{R}^n$ given as

$$F_p(\mathbf{x}, t) = \int_{-r}^{t} f_p(\mathbf{x}, t)\, dt.$$

Then

$$\begin{aligned} d\theta_p &= \tilde{\omega}_p \wedge dt + (-1)^{n-1} f_p(\mathbf{x}, t)\, dt \wedge dx^1 \\ &\wedge \dots \wedge dx^{n-1} = \tilde{\omega}_p \wedge dt + f_p(\mathbf{x}, t)\, dx^1 \\ &\wedge \dots \wedge dx^{n-1} \wedge dt = \omega_p. \end{aligned}$$

The form θ_p is obviously zero outside the bar $\mathring{\mathbf{I}}_r^{n-1} \times \mathbb{R}$. On the other hand, since

$$\int_{-r}^{r} \int_{\mathbb{R}^{n-1}} w_p(\mathbf{x}, t)\, d\mathbf{x}\, dt = \int \omega_p = 0,$$

we have $F_p(\mathbf{x}, t) = 0$ for $t \geqslant r$ (and, of course, for $t \leqslant -r$). Moreover, since $\omega_{t,p} = 0$ for $|t| \geqslant r$, we have $\theta_{t,p} = 0$ for $|t| \geqslant r$, i.e. $\vartheta_p = 0$ for $|t| \geqslant r$. Consequently, $\theta_p = 0$ for $|t| \geqslant r$, i.e. $\theta_p = 0$ outside the cube $\mathring{\mathbf{I}}_r^n$. □

Corollary 1. *If there is*

$$\int_{\mathbb{R}^n} \omega = 0 \tag{8}$$

for a finite form ω of degree n on the space $\mathbb{R}^n$, then there is on $\mathbb{R}^n$ a finite form θ of degree $n-1$ such that

$$\omega = d\theta. \tag{9}$$

If $\omega = 0$ outside $\mathring{\mathbf{I}}_r^n$, then it may be assumed in addition that $\theta = 0$ also outside $\mathring{\mathbf{I}}_r^n$. □

Remark 2. It will be shown in the next lecture that condition (8) is not only sufficient but also necessary for equation (9) to hold.

Now let $\mathscr{X}$ be an arbitrary n-dimensional manifold (Hausdorff and paracompact) and (U, h) be a chart in $\mathscr{X}$. A finite form ω of degree n on $\mathscr{X}$ is said to be concentrater (on U, we mean) if $\omega = 0$ outside U. A form $(h^{-1})^*\omega$ is defined for such a form ω on an open set $h(U) \subset \mathbb{R}^n$. That form is zero outside some compact set $C \subset h(U)$, from which it follows that if it is assumed to be zero outside $h(U)$, then we obtain a smooth (and finite) form on the entire $\mathbb{R}^n$. This form is also denoted by $(h^{-1})^*\omega$.

Let

$$I\omega = \left| \int\limits_{\mathbb{R}^n} (h^{-1})^* \omega \right|.$$

Since the integral may only change sign with a different choice of the diffeomorphism h, the number $I\omega$ depends only on the form ω (is defined correctly).

[If the manifold $\mathscr{X}$ is oriented, which we prefer not to assume for the time being, then the number $I\omega$ is nothing but the absolute value of $\int \omega$.]

A finite concentrated form ω is said to be *essential* if $I\omega \neq 0$, and *inessential* if $I\omega = 0$. An essential form for which $I\omega = 1$ is called *normed.*

A finite form ω on a manifold $\mathscr{X}$ is said to be *finitely cohomologous* to zero if there is a finite form θ on $\mathscr{X}$ such that

$$d\theta = \omega.$$

Lemma 1. *Every concentrated inessential finite form ω is finitely cohomologous to zero.*

Proof. It is clear that we may assume without loss of generality that the set $h(U) \subset \mathbb{R}^n$ is a cube $\mathring{\mathrm{I}}^n$ and that the form $(h^{-1})^*\omega$ (considered as a form on the entire space $\mathbb{R}^n$) is zero outside $\mathring{\mathrm{I}}^n$. Under the hypothesis

$$\int\limits_{\mathbb{R}^n} (h^{-1})^* \omega = 0.$$

It follows that by Corollary 1 to Proposition 2 there is on $\mathbb{R}^n$ a finite form θ which is zero outside $\mathring{\mathrm{I}}^n$, such that

$$(h^{-1})^*\omega = d\theta.$$

Consider a form $h^*\theta$. This form is defined on U and satisfies on U the relation

$$d(h^*\theta) = \omega.$$

Moreover, the form $h^*\theta$ is zero outside some compact subset of a coordinate neighbourhood U, the subset being the image for h^{-1} of a compact subset of the cube $\mathring{\mathrm{I}}^n = h(U)$ outside which the form θ is zero. Assuming $h^*\theta$ to be zero outside U, we therefore obtain on $\mathscr{X}$ a smooth finite differential form θ_1 which satisfies the relation

$$d\theta_1 = \omega$$

both on and outside U, i.e. on the entire $\mathscr{X}$. □

Lemma 2. *For any coordinate neighbourhood $U \subset \mathscr{X}$ there is on $\mathscr{X}$ a normed finite form which is concentrated on U.*

Proof. By Proposition 2 of Lecture 14 there is a smooth nonnegative function φ on $\mathscr{X}$ which is zero outside U and equals unity on some open set $W \subset U$ with compact closure $\overline{W} \subset U$. We define a form $p \mapsto \omega_p$ on $\mathscr{X}$ setting

$$\omega_p = \begin{cases} \varphi(p)\, dx_p^1 \wedge \ldots \wedge dx_p^n & \text{if } p \in U, \\ 0 & \text{otherwise,} \end{cases}$$

where $x^1, \ldots, x^n$ are the local coordinates on U. It is clear that the form ω is smooth, finite, and concentrated on U. In addition,

$$(10) \qquad \int_{h(U)} (h^{-1})^*\omega = \int_{h(U)} (\varphi \circ h^{-1})(\mathbf{x})\, d\mathbf{x} > \int_{h(\overline{W})} (\varphi \circ h^{-1})(\mathbf{x})\, d\mathbf{x} = \int_{h(\overline{W})} d\mathbf{x} > 0,$$

where $h\colon U \to h(U)$ is a coordinate homeomorphism. Hence the form ω is essential. Dividing it by the integral $I\omega$ we obtain a normed form concentrated on U. □

Remark 3. If the manifold $\mathcal{X}$ is oriented and the chart $(U, x^1, \ldots, x^n)$ is positive, then the constructed form ω has the property that

$$\int \omega > 0.$$

Coordinate neighbourhoods U and V in a manifold $\mathcal{X}$ are said to be *chained* if there are in $\mathcal{X}$ coordinate neighbourhoods

$$U_0, U_1, \ldots, U_m$$

such that $U_0 = U$, $U_m = V$ and $U_{i-1} \cap U_i \neq \varnothing$ for any $i = 1, \ldots, m$ (cf. Lecture 20). It is clear (see Lemma 1 of Lecture 20) that *on a connected manifold $\mathcal{X}$ any two coordinate neighbourhoods are chained.*

Finite forms ω_0 and ω_1 on a manifold $\mathcal{X}$ are said to be *finitely cohomologous* if there is a finite form θ such that

$$\omega_1 - \omega_0 = d\theta,$$

i.e. if the form $\omega_1 - \omega_0$ is finitely cohomologous to zero.

Proposition 3. *Let ω_0 and ω_1 be normed concentrated finite forms of degree n on an n-dimensional connected manifold $\mathcal{X}$. Then the form ω_1 is finitely cohomologous to either the form ω_0 or the form $-\omega_0$.*

Proof. Let the form ω_0 be concentrated on a coordinate neighbourhood U_0 and the form ω_1 be concentrated on a coordinate neighbourhood U_1, and let $h_0\colon U_0 \to \mathbb{R}^n$ and $h_1\colon U_1 \to \mathbb{R}^n$ be coordinate mappings such that

$$\int_{\mathbb{R}^n} (h_0^{-1})^* \omega_0 = 1 \text{ and } \int_{\mathbb{R}^n} (h_1^{-1})^* \omega_1 = 1. \tag{11}$$

Case 1. Charts (U_0, h_0) and (U_1, h_1) coincide. Then the finite form $\omega_1 - \omega_0$ is concentrated on U_0 and inessential. It follows that by Lemma 1 the form $\omega_1 - \omega_0$ is finitely cohomologous to zero.

Case 2. Coordinate neighbourhoods U_0 and U_1 coincide. Since integrals (11) may only change sign when coordinate mappings are changed, we have

$$\int_{\mathbb{R}^n} (h_1^{-1})^* \omega_0 = \varepsilon, \text{ where } \varepsilon = \pm 1.$$

According to the foregoing, therefore, the form ω_1 is finitely cohomologous to $\varepsilon\omega_0$.

Case 3. Coordinate neighbourhoods U_0 and U_1 intersect:

$$U_0 \cap U_1 \neq \varnothing.$$

By Lemma 2 there is a normed form ω concentrated on a coordinate neighbourhood $U_0 \cap U_1$ (and hence on the coordinate neighbourhoods U_0 and U_1). Since the normed forms ω_0 and ω are concentrated on U_0, by the foregoing the form ω_0 is finitely cohomologous to either the form ω or the form $-\omega$, and since the normed forms ω and ω_1 are concentrated on U_1, similarly the form ω is finitely cohomologous to either the form ω_1 or the form $-\omega_1$. Hence the form ω_1 is finitely cohomologous to either the form ω_0 or the form $-\omega_0$.

Case 4. Coordinate neighbourhoods U_0 and U_1 are chained. An obvious induction on the length of the chain of coordinate neighbourhoods that joins U_0 and U_1 at once reduces this case to the case we have just considered.

Since in a connected manifold any two coordinate neighbourhoods are chained this proves Proposition 2. □

Corollary 1. *Let ω_0 be an arbitrary essential concentrated form of degree n on an n-dimensional connected manifold $\mathscr{X}$. For any finite form ω of degree n on $\mathscr{X}$ there is then a number c such that the form ω is finitely cohomologous to the form $c\omega_0$.*

Proof. It may be assumed without loss of generality that the form ω_0 is normed.

Since under the hypothesis the manifold $\mathscr{X}$ is Hausdorff and paracompact, there is a numerable covering $\{U_\alpha\}$ on it which consists of coordinate neighbourhoods. Let $\{\eta_\alpha\}$ be a partition of unity subordinate to $\{U_\alpha\}$. For any α the form $\eta_\alpha\omega$ is then concentrated on U_α; hence a number

$$c_\alpha = I\,(\eta_\alpha\omega)$$

is defined.

If $c_\alpha \neq 0$, then like ω_0 the form $c_\alpha^{-1}\eta_\alpha\omega$ is normed and concentrated. It follows from Proposition 3, therefore, that this form is finitely cohomologous to a form $\varepsilon_\alpha\omega_0$, where $\varepsilon_\alpha = \pm 1$. Consequently, the form $\eta_\alpha \omega$ is finitely cohomologous to the form $c_\alpha \varepsilon_\alpha \omega_0$.

This conclusion clearly remains valid for $c_\alpha = 0$ as well, because in this case the concentrated form $\eta_\alpha \omega$ is inessential and hence by Lemma 2 finitely cohomologous to zero.

Thus in the expansion

$$\omega = \sum_\alpha \eta_\alpha \omega$$

(containing due to the finiteness of the form only a finite number of nonzero terms) every term $\eta_\alpha \omega$ is finitely cohomologous to $c_\alpha \varepsilon_\alpha \omega_0$. The form ω is therefore finitely cohomologous to $c\omega_0$, where

$$c = \sum_\alpha c_\alpha \varepsilon_\alpha . \quad \square$$

For any $m \geqslant 0$ we put

$$H^m_{\text{fin}} \mathscr{X} = Z^m_{\text{fin}} \mathscr{X} / B^m_{\text{fin}} \mathscr{X},$$

where $Z^m_{\text{fin}} \mathscr{X}$ is the space of all closed finite forms of degree m on $\mathscr{X}$ and $B^m_{\text{fin}} \mathscr{X}$ is its subspace consisting of differentials of finite forms of degree $m-1$. (Notice that in general there are exact finite forms which are not in $B^m_{\text{fin}} \mathscr{X}$.)

Elements of a quotient space $H^m_{\text{fin}} \mathscr{X}$ are called *finite cohomology classes* of a smooth manifold $\mathscr{X}$.

Theorem 1. *For an arbitrary n-dimensional connected Hausdorff and paracompact manifold $\mathscr{X}$ there is*

$$\dim H^n_{\text{fin}} \mathscr{X} \leqslant 1,$$

i.e. either $H^n_{\text{fin}} \mathscr{X} = 0$ or $H^n_{\text{fin}} \mathscr{X} \simeq \mathbb{R}$.

Proof. By Corollary 1 to Proposition 3 the vector space $H^n_{\text{fin}} \mathscr{X}$ is generated by a finite cohomology class $[\omega_0]$ of an arbitrary concentrated essential form ω_0. If therefore $[\omega_0] = 0$, then $H^n_{\text{fin}} \mathscr{X} = 0$, and if $[\omega_0] \neq 0$, then $H^n_{\text{fin}} \mathscr{X} \simeq \mathbb{R}$. $\square$

When $\mathscr{X}$ is an orientable manifold, a more exact result can be obtained.

Choosing on $\mathscr{X}$ one of the two possible orientations (recall that the manifold $\mathscr{X}$ is assumed to be connected) we can construct for every finite form $\omega \in \Omega^n \mathscr{X}$ its integral

$$\int \omega. \tag{12}$$

In Lecture 27 we shall show (see Corollary 1 to Theorem 1 of Lecture 27) that *if* $\omega \in B^n_{\text{fin}}\mathscr{X}$, *then integral* (12) *is zero*. Therefore integral (12) depends only on a finite cohomology class $[\omega] \in H^n_{\text{fin}}\mathscr{X}$ of the form ω (notice that it is trivially closed), i.e. the correspondence

$$[\omega] \mapsto \int \omega \tag{13}$$

correctly defines some homomorphism $H^n_{\text{fin}}\mathscr{X} \to \mathbb{R}$.

Theorem 2. *For any orientable n-dimensional connected Hausdorff and paracompact manifold* $\mathscr{X}$ *the group* $H^n_{\text{fin}}\mathscr{X}$ *is isomorphic to* $\mathbb{R}$. *The isomorphism is defined by correspondence* (13). (*It depends on the choice of orientation on* $\mathscr{X}$.)

Proof. From Remark 3 we know that there are finite forms ω of degree n on $\mathscr{X}$ for which integral (2) is nonzero. This means that homomorphism (13) is nontrivial.

Since $\dim H^n_{\text{fin}}\mathscr{X} \leqslant 1$, this is possible only for $\dim H^n_{\text{fin}}\mathscr{X} = 1$ and then homomorphism (13) is an isomorphism. □

Thus, for a finite cohomology class $[\omega]$ of a finite (not necessarily concentrated) form ω of degree n to generate the vector space $H^n_{\text{fin}}\mathscr{X}$ (to make up its basis), it is necessary and sufficient that the integral of ω over $\mathscr{X}$ be nonzero:

$$\int \omega \neq 0. \tag{14}$$

For any other finite form ω_1 of degree n on $\mathscr{X}$ there is then an equation

$$[\omega_1] = c\,[\omega],$$

where

$$c = \frac{\int \omega_1}{\int \omega}.$$

Notice that condition (14) is trivially satisfied if the form ω is concentrated and essential.

Corollary 1. *For any orientable n-dimensional connected Hausdorff compact manifold* $\mathscr{X}$ *the group* $H^n\mathscr{X}$ *is isomorphic to* $\mathbb{R}$, *i.e.*

$$h^n\mathscr{X} = 1. \quad \square$$

Remark 4. Giving isomorphism (13) is equivalent to giving in a one-dimensional vector space $H^n_{\mathrm{fin}}\mathscr{X}$ a basis for, and hence some orientation of, that vector space. When an orientation of a manifold $\mathscr{X}$ is changed, that orientation is replaced by the opposite orientation. Hence the *orientations of a connected manifold* $\mathscr{X}$ *can be identified with the orientations of a vector space* $H^n_{\mathrm{fin}}\mathscr{X}$.

Remark 5. It can be shown—try to do it!—that the equation $\dim H^n_{\mathrm{fin}}\mathscr{X} = 1$ characterizes orientable manifolds, i.e. a *connected manifold* $\mathscr{X}$ *is nonorientable if and only if* $H^n_{\mathrm{fin}}\mathscr{X} = 0$.

Problem 4. Compute the group $H^n_{\mathrm{fin}}\mathscr{X}$ for a disconnected manfold $\mathscr{X}$. Compare the result with Proposition 2 of Lecture 20.

Lecture 26

Degree of a smooth proper mapping · The algebraic number of inverse images of a regular value · Invariance of the degree under smooth homotopies · Proof of the drum theorem · Invariance of the degree under any homotopies

---

Definition 1. A mapping $f\colon \mathcal{X} \to \mathcal{Y}$ is said to be *proper* if the inverse image $f^{-1}C$ of every compact set $C \subset \mathcal{Y}$ is compact. It is clear that if a smooth mapping $f\colon \mathcal{X} \to \mathcal{Y}$ is proper, then for any finite form ω on $\mathcal{Y}$ the form $f^*\omega$ on $\mathcal{X}$ is also finite. For any $m \geqslant 0$ therefore every smooth proper mapping $f\colon \mathcal{X} \to \mathcal{Y}$ induces a homomorphism

$$f^*\colon H^m_{\text{fin}}\mathcal{Y} \to H^m_{\text{fin}}\mathcal{X}.$$

We consider this homomorphism in the special case where both manifolds $\mathcal{X}$ and $\mathcal{Y}$ are oriented and have the same dimension $n = m$. We shall also assume that the *manifold $\mathcal{Y}$ is connected.*

Suppose a finite form ω of degree n on $\mathcal{Y}$ is finitely not cohomologous to zero (its finite cohomology class $[\omega]$ makes up a basis for a vector space $H^n_{\text{fin}}\mathcal{Y}$). Then, as we know, $\int\limits_{\mathcal{Y}} \omega \neq 0$, and for any other finite form ω_1 of degree n on $\mathcal{Y}$ there is an equation $[\omega_1] = c\,[\omega]$, where

$$c = \frac{\int\limits_{\mathcal{Y}} \omega_1}{\int\limits_{\mathcal{Y}} \omega}.$$

Since the forms ω_1 and $c\omega$ are finitely cohomologous, the forms $f^*\omega_1$ and $f^*(c\omega) = cf^*\omega$ are also finitely cohomologous, and hence—in the case where $[f^*\omega] \neq 0$ in $H^n_{\text{fin}}\mathscr{X}$—there is an equation

$$c = \frac{\int_{\mathscr{X}} f^*\omega_1}{\int_{\mathscr{X}} f^*\omega}.$$

For $c \neq 0$, i.e. for $\int_{\mathscr{Y}} \omega_1 \neq 0$, it follows that

$$\frac{\int_{\mathscr{X}} f^*\omega}{\int_{\mathscr{Y}} \omega} = \frac{\int_{\mathscr{X}} f^*\omega_1}{\int_{\mathscr{Y}} \omega_1},$$

i.e. the number

$$\deg f = \frac{\int_{\mathscr{X}} f^*\omega}{\int_{\mathscr{Y}} \omega} \tag{1}$$

is independent of the choice of the form ω.

Clearly this conclusion remains valid also for $f^*[\omega] = 0$.

Remark 1. Algebraically the above reasoning rests on the obvious fact that every linear mapping $\mathbb{R} \to \mathbb{R}$ is a multiplication by some number.

Definition 2. Number (1) is the *degree* of a proper smooth mapping $f\colon \mathscr{X} \to \mathscr{Y}$.

Equation (1) may be rewritten as

$$\int_{\mathscr{X}} f^*\omega = \deg f \cdot \int_{\mathscr{Y}} \omega \tag{2}$$

and in this representation it holds for any finite form ω of maximal degree on $\mathscr{Y}$.

Note that for the degree $\deg f$ to be defined it is necessary that the mapping f be proper. This condition is always satisfied when the manifold $\mathscr{X}$ is compact. Thus *if* *the manifold $\mathscr{X}$ is compact, then the degree* deg f *is defined for any smooth mapping* $f\colon \mathscr{X} \to \mathscr{Y}$ (of course, assuming as before that the manifold $\mathscr{Y}$ is connected).

Problem 1. Show that if $\mathscr{X}$ is a compact manifold, and $\mathscr{Y}$ is not, then the degree of any mapping $\mathscr{X} \to \mathscr{Y}$ is zero.

Remark 2. For the degree to be defined, it is also necessary that the manifolds $\mathscr{X}$ and $\mathscr{Y}$ should be not only orientable, but also oriented. When the orientation of one of them is changed, the degree changes sign. For $\mathscr{Y} = \mathscr{X}$, however, a change in the orientation of $\mathscr{X}$ leaves the degree unchanged (the degree changes the sign twice). Therefore one may speak about the degree of proper mappings $\mathscr{X} \to \mathscr{X}$ (where $\mathscr{X}$ is a connected, Hausdorff and paracompact manifold) without assuming $\mathscr{X}$ to be necessarily oriented (it is only necessary for it to be orientable).

By the Sard theorem (see Lecture 15) the smooth mapping $f\colon \mathscr{X} \to \mathscr{Y}$ has regular values $q \in \mathscr{Y}$, and by Proposition 1 of Lecture 13, for any regular value $q \in \mathscr{Y}$ the set $f^{-1}(q)$ is an embedded zero-dimensional submanifold of a manifold $\mathscr{X}$ (recall that under the hypothesis $\dim \mathscr{Y} = \dim \mathscr{X}$), i.e. consists of isolated points (see Lecture 7). Since f is a proper mapping, the set is compact and hence finite.

Let p be an arbitrary point of $f^{-1}(q)$, $(U, h) = (U, x^1, \ldots, x^n)$ be an arbitrary chart of $\mathscr{X}$ centred at p, $(V, k) = (V, y^1, \ldots, y^n)$ be a positive chart of $\mathscr{Y}$ centred at a point q, and

$$y^j = f^j(x^1, \ldots, x^n), \quad j = 1, \ldots, n, \tag{3}$$

be functions expressing in the charts the mapping f. Under the hypothesis the Jacobian

$$D_f(p) = \det \left\| \frac{\partial f^j}{\partial x^i} \right\|, \quad i, j = 1, \ldots, n,$$

of functions (3) is nonzero at p. Therefore $\operatorname{sgn} D_f(p) = \pm 1$ is defined. Clearly this sign is independent of the choice of (positive!) charts (U, h) and (V, k). We shall call it the *sign of the mapping* f at p and denote by $\operatorname{sgn}_p f$.

Let

$$\sum_{f(p)=q} \operatorname{sgn}_p f \tag{4}$$

be the sum of the signs of f over all points $p \in f^{-1}(q)$ (the "algebraic number of the inverse images of the point q under f"). This sum is defined, since $f^{-1}(q)$ is a finite set.

Proposition 1. *Number* (4) *is independent of the choice of a point* $q \in \mathscr{Y}$ *and equals the degree* $\deg f$ *of a mapping* f.

Proof. Let W be an arbitrary neighbourhood of the point q in $\mathscr{Y}$ whose closure $\overline{W}$ is compact.

Since the mapping f is continuous and the set of all its critical points is closed, every point $p \in f^{-1}(q)$ has a neighbourhood $U'_p \subset f^{-1}W$ containing no critical points of f. Of course, neighbourhoods U'_p may be chosen so that they should not intersect for distinct points $p \in f^{-1}(q)$, all the sets $f(U'_p)$ should be open in $\mathscr{Y}$ and all the mappings $f \mid U'_p : U'_p \to f(U'_p)$ should be diffeomorphisms.

Let

$$U' = \bigcup_{f(p)=q} U'_p$$

be the union of all neighbourhoods U'_p and let $C = f^{-1}\overline{W} \setminus U'$. The set C is compact and has the property that $q \notin f(C)$. Its image $f(C)$ is also compact and hence closed. Consequently, there is a coordinate neighbourhood V' of a point q on a manifold $\mathscr{Y}$ such that $V' \subset W \setminus f(C)$ and hence such that $f^{-1}V' \subset U'$.

Let

$$V = \bigcap_{f(p)=q} f(U'_p) \cap V'.$$

Since the set $f^{-1}(q)$ is finite, the set V (containing the point q) is open and hence it is a neighbourhood of q. For any point $p \in f^{-1}(q)$ we put

$$U_p = U'_p \cap f^{-1}V.$$

Obviously the sets U_p have the following properties:

(a) Every set U_p is a neighbourhood of a point p.

(b) Neighbourhoods U_p corresponding to distinct points p do not intersect.

(c) The union U of all neighbourhoods U_p, $p \in f^{-1}(q)$, is the inverse image $f^{-1}V$ of a neighbourhood V.

(d) On every neighbourhood U_p the mapping f is its diffeomorphism on the neighbourhood V.

Since the neighbourhood V is contained in the coordinate neighbourhood V', it is itself a coordinate neighbourhood, i.e. the support of some chart $(V, k) = (V, y^1, \ldots, y^n)$ of $\mathcal{Y}$ which may be assumed without loss of generality to be positive. By virtue of (d) it follows that for any point $p \in f^{-1}(q)$ the pair (U_p, h_p), where $h_p = k \circ (f|_{U_p})$, is also a chart. Since in charts (U_p, h_p) and (V, k) the mapping f can be written by formulas of the form $y^i = x^i$, $i = 1, \ldots, n$, the chart (U_p, h_p) is positive if and only if $\operatorname{sgn}_p [f] = 1$.

As we know, to compute $\deg f$ we may use an arbitrary finite form ω, finitely not cohomologous to zero, of degree n on $\mathcal{Y}$. Making use of this freedom, we take as ω an essential form concentrated on V. Such a form exists according to Lemma 2 of Lecture 25.

But if a form ω is concentrated on V, then the form $f^*\omega$ is zero outside neighbourhoods U_p, and therefore

$$\int_{\mathcal{X}} f^*\omega = \sum_{f(p)=q} \int_{U_p} f^*\omega,$$

where on the right U_p is considered to be an open submanifold provided with an orientation induced by the orientation of the manifold $\mathcal{X}$. If therefore

$$\omega = w\, dy^1 \wedge \ldots \wedge dy^n \text{ on } V,$$

and hence

$$f^*\omega = (w \circ f)\, dx^1 \wedge \ldots \wedge dx^n \text{ on } U_p,$$

where $x^1 = y^1 \circ f, \ldots, x^n = y^n \circ f$, then

$$\int_{U_p} f^*\omega = \varepsilon_p \int_{h_p(U_p)} (w \circ f \circ h_p^{-1})(\mathbf{x})\, d\mathbf{x},$$

where $\varepsilon_p = 1$ if (U_p, h_p) is a positive chart and $\varepsilon_p = -1$ otherwise. On the other hand, since $h_p = k \circ (f \mid U_p)$ and hence $h_p(U_p) = k(V)$ and $w \circ f \circ h_p^{-1} = w \circ k^{-1}$,

we have

$$\int_{h_p(U_p)} (w \circ f \circ h_p^{-1})(\mathbf{x})(d\mathbf{x}) = \int_{k(V)} (w \circ k^{-1})(\mathbf{x})\, d\mathbf{x} = \int_{\mathscr{Y}} \omega.$$

Moreover, according to the above remark the equation $\varepsilon_p = +1$ holds if and only if $\operatorname{sgn}_p f = 1$. Since $\varepsilon_p = \pm 1$ and $\operatorname{sgn}_p f = \pm 1$, this means that

$$\varepsilon_p = \operatorname{sgn}_p f \text{ for any point } p \in f^{-1}(q).$$

Comparing all these facts we immediately get

$$\int_{\mathscr{X}} f^*\omega = \left(\sum_{f(p)=q} \operatorname{sgn}_p f \right) \int_{\mathscr{Y}} \omega.$$

By definition (see formula (1)) it is this that implies that number (4) is equal to deg f of the mapping f. □

Corollary 1. *The degree* deg f *of an arbitrary proper smooth mapping* $f\colon \mathscr{X} \to \mathscr{Y}$ *is an integer.* □

Corollary 2. *If a mapping* $f\colon \mathscr{X} \to \mathscr{Y}$ *is not subjective, then* $\deg f = 0$.

Proof. It suffices to notice that any point $q \notin f(\mathscr{X})$ is a regular value of the mapping f. □

Definition 3. Continuous (proper) mappings $f, g\colon \mathscr{X} \to \mathscr{Y}$ are said to be (*properly*) *homotopic* if there is a continuous (proper) mapping $F\colon \mathscr{X} \times [0, 1] \to \mathscr{Y}$ such that

$$F(p, 0) = f(p), \quad F(p, 1) = g(p)$$

for any point $p \in \mathscr{X}$.

The mapping F is called a (*proper*) *homotopy* connecting the mappings f and g. It is convenient to identify it with a family $\{f_t\}$ of continuous (proper) mappings $f_t\colon \mathscr{X} \to \mathscr{Y}$, $0 \leqslant t \leqslant 1$, defined as follows:

$$f_t(p) = F(p, t), \quad p \in \mathscr{X}.$$

A homotopy F is said to be *smooth* if it is a restriction to $\mathscr{X} \times [0, 1]$ of some smooth mapping $\mathscr{X} \times (a, b) \to \mathscr{Y}$, where (a, b) is an interval of the axis $\mathbb{R}$ that contains a closed interval $[0, 1]$. Smooth (and proper) mappings $f, g\colon \mathscr{X} \to \mathscr{Y}$ connected by a smooth (and proper) homotopy $F\colon \mathscr{X} \times [0, 1] \to \mathscr{Y}$ are called *smoothly* (and *properly*) *homotopic.*

Since for a smooth and proper homotopy $F\colon \mathscr{X} \times [0, 1] \to \mathscr{Y}$ all mappings $f_t\colon \mathscr{X} \to \mathscr{Y}$, $0 \leqslant t \leqslant 1$, are smooth and proper, for every one of them a degree

$$\deg f_t = \frac{\int\limits_{\mathscr{X}} f_t^*\omega}{\int\limits_{\mathscr{Y}} \omega}$$

is defined.

The form $f_t^*\omega$ involved in this formula is obviously continuously (even smoothly) dependent on t, i.e. in every chart its only coefficient is a continuous function of t (and of local coordinates, of course). Therefore (see Problem 3 of Lecture 25) the number $\deg f_t$ is also continuously dependent on t. Hence, being an integer it is constant. In particular, $\deg f = \deg f_0 = \deg f_1 = \deg g$.

Thus, if smooth proper mappings f, $g\colon \mathscr{X} \to \mathscr{Y}$ are smoothly and properly homotopic, then their degrees are equal:

$$\deg f = \deg g. \tag{5}$$

This statement is known as the **theorem on the homotopic invariance of the degree.**

Now we can easily prove the drum theorem advertised in Lecture 9.

Proof of Theorem 1 of Lecture 9. Suppose there is a retraction

$$r\colon \mathbb{B}^n \to \mathbb{S}^{n-1}, \quad n \geqslant 2,$$

and let $r(0) = \mathbf{s}_0$. Then the formula

$$F(\mathbf{x}, t) = r(t\mathbf{x}), \ \mathbf{x} \in \mathbb{S}^{n-1}, \ 0 \leqslant t \leqslant 1, \tag{6}$$

defines a homotopy

$$F\colon \mathbb{S}^{n-1} \times [0, 1] \to \mathbb{S}^{n-1}$$

connecting a constant mapping

$$\text{const}\colon \mathbb{S}^{n-1} \to \mathbb{S}^{n-1}, \quad \mathbf{x} \mapsto \mathbf{s}_0$$

with an identity mapping id: $\mathbb{S}^{n-1} \to \mathbb{S}^{n-1}$, $\mathbf{x} \mapsto \mathbf{x}$. By virtue of the homotopic invariance of the degree mappings const and id (obviously smooth) must therefore have the same degree (that these mappings should be proper need not bother us, since the sphere $\mathbb{S}^{n-1}$ is compact). But it is clear that deg id $= 1$ and deg const $= 0$. This contradiction proves that the retraction r cannot exist. □

Remark 3. This proof fails for $n = 1$ (if only because the sphere $\mathbb{S}^0$ consists of two points and is not a connected manifold). But in this case it is obvious that the retraction $\mathbb{B}^1 \to \mathbb{S}^0$ cannot exist (the connected segment $\mathbb{B}^1$ cannot be continuously mapped onto the disconnected sphere $\mathbb{S}^0$).

An attentive reader must have noticed that the above proof of the drum theorem has a gap, and therefore, strictly speaking, cannot be considered to be a proof. Indeed, homotopy (6) is in general only continuous, whereas equation (5) assumes that the homotopy connecting mappings f and g is smooth. To lay a firm foundation for the drum theorem we must therefore prove (at least for mappings $\mathbb{S}^{n-1} \to \mathbb{S}^{n-1}$) that *equation* (5) *remains valid also when the homotopy connecting the mappings f and g is only continuous.* To do this it is sufficient, of course, to prove the following general proposition.

Proposition 2. *For any compact* (*and Hausdorff*) *manifolds $\mathscr{X}$ and $\mathscr{Y}$ smooth mappings f, g: $\mathscr{X} \to \mathscr{Y}$ are homotopic if and only if they are smoothly homotopic.*

[By making the technicalities more involved a similar statement can be proved for proper mappings of noncompact manifolds as well, but since we need only the case where $\mathscr{X} = \mathscr{Y} = \mathbb{S}^{n-1}$ to prove the drum theorem, we shall not do this. Besides, strictly speaking, we prove Proposition 2 only under some additional assumptions imposed on $\mathscr{X}$ and $\mathscr{Y}$ which are trivially satisfied for $\mathscr{X} = \mathscr{Y} = \mathbb{S}^{n-1}$. (In fact they are satisfied for any $\mathscr{X}$ and $\mathscr{Y}$, but we shall be able to establish this result only in the next semester.)]

Note that Proposition 2 does not assume the dimensions of manifolds $\mathscr{X}$ and $\mathscr{Y}$ to be equal (and the manifolds $\mathscr{X}$ and $\mathscr{Y}$ themselves may be disconnected).

Manifolds $\mathscr{X}$ and $\mathscr{Y}$ will be assumed to be embedded in the space $\mathbb{R}^n$, where n is some sufficiently large number.

(Thus, here we depart from the custom of using n to denote the dimension of a manifold $\mathscr{X}$.) According to the embedding theorem (Proposition 1 of Lecture 14) this may be assumed without any loss of generality.

Definition 4. A subset $\mathscr{X}$ of $\mathbb{R}^n$ is said to be a *neighbourhood retract* if there is an open set $O \supset \mathscr{X}$ that is a *retraction* onto $\mathscr{X}$. When $\mathscr{X}$ is a submanifold, it is in addition required that there should be a retraction $r: O \to \mathscr{X}$ which is a smooth mapping.

For example, a *neighbourhood retract is the sphere* S^{n-1} *of* $\mathbb{R}^n$. (In this case we may take as a neighbourhood O a punctured space $\mathbb{R}^n \setminus \{0\}$ and define the retraction r: $O \to S^{n-1}$ as follows: $r(\mathbf{x}) = \dfrac{\mathbf{x}}{|\mathbf{x}|}$, $\mathbf{x} \in \mathbb{R}^n \setminus \{0\}$.)

We prove Proposition 2 only on the assumption that *submanifolds* $\mathscr{X}$ *and* $\mathscr{Y}$ *of* $\mathbb{R}^n$ *are neighbourhood retracts.* [It will be shown in the next semester that this assumption may be made without any loss of generality. Moreover, at all events it is satisfied when $\mathscr{X} = \mathscr{Y} = S^{n-1}$.]

Proof of Proposition 2. Of course, if smooth mappings f, g: $\mathscr{X} \to \mathscr{Y}$ are smoothly homotopic, then they are also homotopic. Therefore it is only necessary for us to prove the converse. The natural way is to transform an arbitrary continuous homotopy F: $\mathscr{X} \times [0, 1] \to \mathscr{Y}$ connecting the mappings f and g into a smooth one.

Let r: $O_{\mathscr{X}} \to \mathscr{X}$ be a retraction onto $\mathscr{X}$ of some neighbourhood $O_{\mathscr{X}} \supset \mathscr{X}$. We define

$$F': O_{\mathscr{X}} \times \mathbb{R} \to \mathscr{Y}$$

as

$$F'(\mathbf{x}, t) = \begin{cases} f(r(\mathbf{x})) & \text{if } t \leqslant 3/7 \\ F(r(\mathbf{x}), 7t-3) & \text{if } 3/7 \leqslant t \leqslant 4/7, \\ g(r(\mathbf{x})) & \text{if } 4/7 \leqslant t. \end{cases}$$

Clearly, the mapping F' is continuous (even smooth for $t < 3/7$ and $t > 4/7$) and its restriction to $\mathscr{X} \times [0, 1]$ is a homotopy connecting the mappings f and g.

The following theorem is known in calculus.

Weierstrass theorem (polynomial approximation). *Let O be an open set of $\mathbb{R}^n$, K be a compact subset of O, and F: $O \to \mathbb{R}^m$ be a continuous mapping. For any $\varepsilon > 0$ there* *is then a polynomial mapping* $P\colon \mathbb{R}^n \to \mathbb{R}^m$ *(given by the polynomial functions of coordinates) such that*

$$|P(\mathbf{x}) - F(\mathbf{x})| < \varepsilon \text{ for any point } \mathbf{x} \in K.$$

[Notice that it is obviously enough to prove the theorem for $m = 1$, i.e. when F is a number function. It is only this case that is usually treated in calculus.]

We apply the Weierstrass theorem to an open set $O_{\mathscr{X}} \times \mathbb{R}$ (considered to be a subset of the space $\mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R}$), to a compact set $\mathscr{X} \times [0, 1]$ and to a mapping F' (considered to be a mapping in $\mathbb{R}^n$). Again denoting the restriction of a polynomial mapping P to $\mathscr{X} \times [0, 1]$ by P we obtain by the theorem that *for any* $\varepsilon > 0$ *there is a smooth mapping*

$$P\colon \mathscr{X} \times [0, 1] \to \mathbb{R}^n$$

such that

$$|P(\mathbf{x}, t) - F'(\mathbf{x}, t)| < \varepsilon$$
$$\text{for any point } (\mathbf{x}, t) \in \mathscr{X} \times [0, 1].$$

Now let λ be a smooth function $\mathbb{R} \to \mathbb{R}$ such that

$$\lambda(t) = \begin{cases} 0 & \text{if } t \leqslant 1/7 \text{ or } 6/7 \leqslant t, \\ 1 & \text{if } 2/7 \leqslant t \leqslant 5/7 \end{cases}$$

and $0 \leqslant \lambda(t) \leqslant 1$ for any $t \in \mathbb{R}$. [Such a function can be given, for example, as follows:

$$\lambda(t) = \frac{\alpha\left(\frac{5}{14} - \left|\frac{1}{2} + t\right|\right)}{\alpha\left(\frac{5}{14} - \left|\frac{1}{2} + t\right|\right) + \alpha\left(\left|t + \frac{1}{2}\right| - \frac{3}{14}\right)},$$

$$-\infty < t < +\infty,$$

where α is the function in Lemma 1 of Lecture 1; cf. Remark 3 of Lecture 1.] We put

$$G(\mathbf{x}, t) = F'(\mathbf{x}, t) + \lambda(t)(P(\mathbf{x}, t) - F'(\mathbf{x}, t))$$

for any point $(\mathbf{x}, t) \in \mathscr{X} \times [0, 1]$.

Since $\lambda(t) = 0$ for $t = 0$ and $t = 1$, we have $G(\mathbf{x}, 0) = F'(\mathbf{x}, 0) = f(\mathbf{x})$ and $G(\mathbf{x}, 1) = F'(\mathbf{x}, 1) = g(\mathbf{x})$

for any point $\mathbf{x} \in \mathscr{X}$, i.e. G is a homotopy connecting mappings f and g. Since for $t < 3/7$ and $4/7 < t$ the mapping F' (and hence the mapping G) is smooth, and for $2/7 < t < 5/7$ the mapping G coincides with the mapping P (and is therefore also smooth), G is a smooth homotopy.

We have constructed a smooth homotopy, but it takes on values not in the manifold $\mathscr{Y}$ but in $\mathbb{R}^n$. To improve the situation we recall that the construction of a homotopy G depended on the parameter $\varepsilon > 0$ and that

$$|G(\mathbf{x}, t) - F'(\mathbf{x}, t)| \leqslant \lambda(t)\, |P(\mathbf{x}, t) - F'(\mathbf{x}, t)| < \varepsilon$$

for any point $(\mathbf{x}, t) \in \mathscr{X} \times [0, 1]$. Since $F'(\mathbf{x}, t) \in \mathscr{Y}$, this means by definition that the *distance of the point* $G(\mathbf{x}, t)$ *from the manifold* $\mathscr{Y}$ *is less than* ε, i.e. the point $G(\mathbf{x}, t)$ is contained in an ε-neighbourhood of $\mathscr{Y}$.

On the other hand, there is under the hypothesis a neighbourhood of the manifold $O_{\mathscr{Y}}$ that is a retract onto $\mathscr{Y}$. Since $\mathscr{Y}$ is a compact manifold, there is $\varepsilon > 0$ such that the entire ε-neighbourhood of $\mathscr{Y}$ is contained in $O_{\mathscr{Y}}$ (prove this!). Consequently, the homotopy G constructed for this ε has the property that $G(\mathbf{x}, t) \in O_{\mathscr{Y}}$ for any point $(\mathbf{x}, t) \in \mathscr{X} \times [0, 1]$. Therefore the formula

$$H(\mathbf{x}, t) = r_{\mathscr{Y}}(G(\mathbf{x}, t)), \quad (\mathbf{x}, t) \in \mathscr{X} \times [0, 1],$$

where $r_{\mathscr{Y}}\colon O_{\mathscr{Y}} \to \mathscr{Y}$ is a smooth retraction, defines a smooth homotopy $H\colon \mathscr{X} \times [0, 1] \to \mathscr{Y}$ connecting mapping $r_{\mathscr{Y}} \circ f = f$ with mapping $r_{\mathscr{Y}} \circ g = g$.

This completes the proof (on the assumption that the manifolds $\mathscr{X}$ and $\mathscr{Y}$ are neighbourhood retracts) of Proposition 2. □

The proof of the drum theorem is thus also complete.

Remark 4. Just as we have approximated an arbitrary homotopy $\mathscr{X} \times [0, 1] \to \mathscr{Y}$ by a smooth homotopy, it is possible to approximate any continuous mappings $f\colon \mathscr{X} \to \mathscr{Y}$ by a smooth mapping.

Problem 2. In the next semester we shall prove that for any compact manifold $\mathscr{Y}$ embedded in $\mathbb{R}^n$ there is a constant $d > 0$ such that any two points $p, q \in \mathscr{Y}$ the distance between which (measured by $\mathscr{Y}$) is less than d can be joined in $\mathscr{Y}$ by a unique shortest path (for a sphere the constant d is π and the shortest path is an arc of the great circle). Making use of this show that *any two sufficiently close mappings* $\mathscr{X} \to \mathscr{Y}$ *are homotopic.*

In particular, it follows that any two smooth mappings $\mathscr{X} \to \mathscr{Y}$ which approximate sufficiently closely a given continuous mapping $f\colon \mathscr{X} \to \mathscr{Y}$ are of the same degree (if the manifolds $\mathscr{X}$ and $\mathscr{Y}$ are oriented, compact and have the same dimension). That degree is called the *degree of a continuous mapping* f.

Problem 3. Prove that the *degrees of homotopic continuous mappings are equal.*

Lecture 27

Domains with regular boundaries · Stokes' theorem · Gauss-Ostrogradskii, Green's, and Newton-Leibniz formulas · Manifolds with border sets · Interior and boundary points · Embedded ∂-submanifolds · Stokes' theorem for manifolds with border set and for ∂-submanifolds · Stokes' theorem for surface integrals · Stokes' theorem for singular submanifolds · Line integrals of the second kind

---

Let $\mathscr{X}$ be as before a smooth n-dimensional Hausdorff and paracompact manifold.

Definition 1. A subset D of a manifold $\mathscr{X}$ is said to be a *domain with regular boundary* if

1° the subset D is the closure of its own interior

$$D = \overline{\operatorname{Int} D};$$

2° its boundary (or frontier)

$$\operatorname{Fr} D = \overline{D} \setminus \operatorname{Int} D$$

is an embedded $(n-1)$-dimensional submanifold.

For such a domain D the boundary Fr D is called its *border set* and denoted by ∂D.

As we know (see Lecture 13), for every point of the submanifold ∂D in the manifold $\mathscr{X}$ there is a chart $(U, h) = (U, x^1, \ldots, x^n)$ containing that point such that the pair

$$(V, k) = (V, y^1, \ldots, y^{n-1}),$$

where $V = U \cap \partial D$ and $y^1 = x^2|_V, \ldots, y^{n-1} = x^n|_V$, is a chart on ∂D and the equation $x^1(p) = 0$ holds for a point $p \in U$ if and only if $p \in \partial D$. We may assume without loss of generality that $x^1 < 0$ on $U \cap \operatorname{Int} D$.

We shall say that a chart (U, h) having these properties is *adapted to D* and that a chart (V, k) is *cut out* (or *excised*) *on* ∂D by the chart (U, h).

Now let $(U, h) = (U, x^1, \ldots, x^n)$ and $(U', h') = (U', x^{1'}, \ldots, x^{n'})$ be two charts on $\mathcal{X}$ adapted to D and let (V, k) and (V', k') be the charts they cut out on ∂D. hince the function $x^{1'}$ is identically zero on $V \cap V'$, we Save on $V \cap V'$

$$\frac{\partial x^{1'}}{\partial x^k} = 0, \quad k = 2, \ldots, n.$$

Therefore

$$\det \frac{\partial h'}{\partial h} = \frac{\partial x^{1'}}{\partial x^1} \cdot \det \frac{\partial k'}{\partial k} \quad \text{on } V \cap V'.$$

On the other hand, since $x^{1'} < 0$ if and only if $x^1 < 0$, then

$$\frac{\partial x^{1'}}{\partial x^1} > 0 \quad \text{on} \quad V \cap V'.$$

Consequently, *if the charts* (U, h) *and* (U', h') *are positively oriented, then (for* $n > 1$) *the charts* (V, k) *and* (V', k') *are also positively oriented.*

Now suppose that the manifold $\mathcal{X}$ is orientable and oriented. Since for any point $p \in \partial D$, with $n > 1$, there is obviously a positive chart containing the point p and adapted to D, and all positive charts adapted to D cut out on ∂D an atlas of positively oriented charts. The orientation on ∂D given by that atlas is said to be *induced* by the orientation of the manifold $\mathcal{X}$.

When $n = 1$ the domain D is a system of segments (on a straight line or circle) and ∂D consists of their ends. The orientation on the manifold $\mathcal{X}$ gives on those segments a direction, and we introduce on ∂D an orientation (in the sense of Remark 1 of Lecture 25), assuming that the right end of every segment has a plus sign and the left end has a minus sign.

Thus, *in an orientable (oriented) manifold $\mathscr{X}$ the border set of any domain with regular boundary is an orientable (oriented) manifold.*

Since by the Sard theorem (see Lecture 15) the border set of an arbitrary domain with regular boundary is a null set, every compact domain D with regular boundary is cubable. Therefore, on the assumption that the manifold $\mathscr{X}$ is oriented, for any form ω of degree n on $\mathscr{X}$ and any compact domain D with regular boundary an integral

$$\int_D \omega$$

is defined. This integral is defined for noncompact domains D as well provided the form ω is finite.

In particular, for any form ω of degree $n-1$ (finite if D is a noncompact domain) an integral

$$\int_D d\omega \tag{1}$$

is defined. On the other hand, also defined (with respect to the induced orientation of the manifold ∂D) is an integral

$$\int_{\partial D} i^*\omega,$$

where $i\colon \partial D \to \mathscr{X}$ is an embedding. For brevity we shall denote this integral by

$$\int_{\partial D} \omega. \tag{2}$$

Theorem 1 (Stokes' theorem for domains with regular boundaries). *For any domain D with regular boundary of a Hausdorff paracompact oriented n-dimensional manifold $\mathscr{X}$ and any form $\omega \in \Omega^{n-1}\mathscr{X}$ (finite if D is noncompact) there is an equation*

$$\int_D d\omega = \int_{\partial D} \omega. \tag{3}$$

Proof. Clearly charts (U_α, h_α) of the manifold $\mathscr{X}$ either not overlapping with ∂D or adapted to D make up an atlas on $\mathscr{X}$. Since under the hypothesis the manifold $\mathscr{X}$ is paracompact and Hausdorff, there is a partition of unity $\{\eta_\alpha\}$ subordinate to a covering $\{U_\alpha\}$. Since $\omega = \sum\limits_\alpha \eta_\alpha \omega$ and $d\omega = \sum\limits_\alpha d(\eta_\alpha\omega)$, we have

$$\int\limits_D d\omega = \sum_\alpha \int\limits_D d(\eta_\alpha\omega)$$

and

$$\int\limits_{\partial D} \omega = \sum_\alpha \int\limits_{\partial D} \eta_\alpha\omega$$

(clearly the integral is also additive with respect to infinite sums of the type we are considering here having in the neighbourhood of any point just a finite number of nonzero terms). It suffices therefore to prove formula (3) only for forms $\eta_\alpha\omega$, i.e. on the assumption that $\omega = 0$ outside some positive chart $(U, h) = (U, x^1, \ldots, x^n)$ of $\mathscr{X}$, either not overlapping with ∂D or adapted to D. If on U

$$\omega = \sum_{k=1}^{n} (-1)^{k-1} w_k \, dx^1 \wedge \ldots \wedge \widehat{\quad}^{k} \wedge \ldots \wedge dx^n,$$

then

$$d\omega = \left(\sum_{k=1}^{n} \frac{\partial w_k}{\partial x^k} \right) dx^1 \wedge \ldots \wedge dx^n$$

on U (see formula (4) of Lecture 19), and hence

$$\int\limits_D d\omega = \int\limits_{U \cap D} d\omega = \int\limits_{h(U \cap D)} \left(\sum_{k=1}^{n} \frac{\partial w_k}{\partial x^k} \right) dx^1 \ldots dx^n.$$

Functions $w_1, \ldots, w_n$ considered as functions on $h(U)$ are zero outside some closed set, and if therefore they are assumed to be zero outside $h(U)$ on the entire $\mathbb{R}^n$, then again smooth functions result. On the other hand, we may assume without loss of generality that the open set $h(U)$ is bounded, i.e. contained in some cube

$$I_R^n=\{x\in\mathbb{R}^n, \quad |x^1|\leqslant R,\ \ldots,\ \ |x^n|\leqslant R\}$$

(on whose boundary all functions w_k are therefore zero).

Case 1. The chart (U, h) does not overlap with ∂D. In this case integral (2) is obviously zero (since $i^*\omega = 0$), and hence to prove (3) it suffices to establish that integral (1) is zero. We may assume without loss of generality that either $U\subset \mathscr{X}\setminus D$ or $U\subset D$. But for $U\subset\mathscr{X}\setminus D$ integral (1) is trivially zero, and for $U\subset D$ it is expressed by the formula

$$\int_D d\omega=\int_{I_R^n}\left(\sum_{k=1}^{n}\frac{\partial w_k}{\partial x^k}\right)dx^1\ldots dx^n$$

$$=\sum_{k=1}^{n}\int_{-R}^{R}\cdots\int_{-R}^{R}\frac{\partial w_k}{\partial x^k}\,dx^1\ldots dx^n,$$

and hence since for any $k = 1, \ldots, n$ the integral

$$\int_{-R}^{R}\frac{\partial w_k}{\partial x^k}\,dx^k=w_k(x^1, \ldots, R, \ldots, x^n)$$

$$-w_k(x^1, \ldots, -R, \ldots, x^n)$$

is zero, D is also zero.

Case 2. The chart (U, h) is adapted to D. In this case by similar considerations

$$\int_D d\omega=\sum_{k=1}^{n}\int_{-R}^{0}\int_{-R}^{R}\cdots\int_{-R}^{R}\frac{\partial w_k}{\partial x^k}\,dx^1\ldots dx^k$$

$$=\int_{-R}^{R}\cdots\int_{-R}^{R}w_1(0,\ x^2, \ldots, x^n)\,dx^2\ldots dx^n.$$

On the other hand, since $i^*(dx^1) = 0$ and $i^*(dx^2) = dy^1, \ldots, i^*(dx^n) = dy^{n-1}$ (because $x^1 = 0$ and $x^2 = y^1, \ldots, x^n = y^{n-1}$ on ∂D), we have

$i^*\omega = w_1(0,\ y^1, \ldots, y^{n-1})\,dy^1\wedge\ldots\wedge dy^{n-1}$ on $U\cap\partial D$.

Therefore

$$\int_{-R}^{R} \dots \int_{-R}^{R} w_1(0, x^2, \dots, x^n)\, dx^2 \dots dx^n$$

$$= \int_{-R}^{R} \dots \int_{-R}^{R} w_1(0, y^1, \dots, y^{n-1})\, dy^1 \dots dy^{n-1}$$

$$= \int_{U \cap \partial D} i^*\omega = \int_{\partial D} \omega.$$

Hence

$$\int_{D} d\omega = \int_{\partial D} \omega. \ \square$$

Corollary 1. *For any finite form* ω *of degree* $n-1$ *on an oriented n-dimensional paracompact Hausdorff manifold* $\mathscr{X}$ *there is an equation*

$$\int_{\mathscr{X}} d\omega = 0. \tag{4}$$

Proof. Let p_0 be an arbitrary point of $\mathscr{X}$ and let (U, h) be a chart in $\mathscr{X}$ such that $p_0 \in U$ and $h(U) = \mathring{\mathbb{B}}_2^n$, where as before $\mathring{\mathbb{B}}_2^n$ is an open ball of $\mathbb{R}^n$ of radius 2 with centre at the point 0. Also let $D_1 = h^{-1}(\mathbb{B}_1^n)$, where $\mathbb{B}_1^n$ is a closed concentric ball of radius 1, and let $D_2 = \mathscr{X} \setminus \operatorname{Int} D_1$. Clearly both sets D_1 and D_2 are domains with the same regular boundary (border set) $h^{-1}(\mathbb{S}^{n-1})$. (It is said that D_2 is obtained from $\mathscr{X}$ by *drilling out a ball* D_1.) The orientations induced on the border set $h^{-1}(\mathbb{S}^{n-1})$ by the orientation of $\mathscr{X}$ are easily seen to be opposite (which can be written as $\partial D_2 = -\partial D_1$), and hence

$$\int_{\partial D_2} \omega = -\int_{\partial D_1} \omega.$$

Consequently, applying Theorem 1 to the domains D_1 and D_2 and considering that

$$\int\limits_{\mathscr{X}} d\omega = \int\limits_{D_1} d\omega + \int\limits_{D_2} d\omega$$

we immediately obtain (4). □

Formula (4) may be considered to be a special case of the general formula (3) if we agree that the integral of an arbitrary form over an empty set is equal to zero.

We have already made use of Corollary 1 in Lecture 25 (see p. 434).

In the special case where $\mathscr{X}$ is a space $\mathbb{R}^3$ with coordinates x, y, z, every form ω of degree 3 on $\mathscr{X}$ is as follows: $f\,dx \wedge dy \wedge dz$, where f is some function, and for any cubable set $D \subset \mathbb{R}^3$ the integral $\int\limits_D \omega$ is equal to the Riemann integral $\int\limits_D f(\mathbf{x})\,d\mathbf{x}$ which in the case under consideration is usually denoted by

$$\iiint\limits_D f\,dx\,dy\,dz$$

to emphasize three-dimensionality and the possibility of being reduced to the triple Riemann integral on a straight line. By analogy, for any oriented surface $\mathscr{X}$ in $\mathbb{R}^3$ (oriented two-dimensional submanifold) and any form

$$\omega = P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy$$

on $\mathbb{R}^3$ (note the order of the differentials in the second term!) the integral $\int\limits_{\mathscr{X}} \omega$ (more exactly, the integral $\int\limits_{\mathscr{X}} i^*\omega$, where i is an embedding $\mathscr{X} \to \mathbb{R}^3$) is denoted by

$$\iint\limits_{\mathscr{X}} P\,dy\,dz + Q\,dz\,dx + R\,dx\,dz. \tag{5}$$

Since

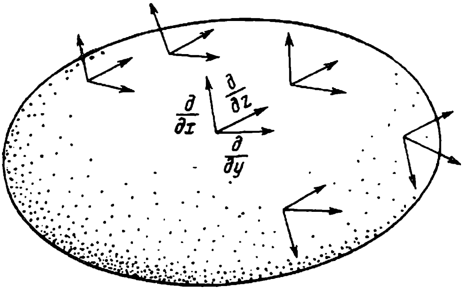
$$d\omega = \left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right) dx \wedge dy \wedge dz,$$

we obtain as a special case of Theorem 1 that *for any domain $D \subset \mathbb{R}^3$ with regular boundary and any functions P, Q, and R we have*

$$\iiint_D \left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right) dx\, dy\, dz = \iint_{\partial D} P\, dy\, dz + Q\, dz\, dx + R\, dx\, dy. \tag{6}$$

Formula (6) is called the **Gauss-Ostrogradskii formula.**

Remark 1. In (6) all integrals are assumed to exist, i.e. either the functions P, Q, and R are finite or the domain


$\frac{\partial}{\partial x}$
$\frac{\partial}{\partial z}$
$\frac{\partial}{\partial y}$

D is compact (and the functions P, Q, and R are defined on D, of course). At the same time, for this formula to be true there is no need to necessarily assume that D is a domain with regular boundary; it suffices to, say, regard the boundary of D as *piecewise regular* (in the obvious sense).

Remark 2. The orientation of a plane in $\mathbb{R}^3$ is given by its side, i.e. by a vector orthogonal to the plane. Therefore the orientation of a surface in $\mathbb{R}^3$ is given by a field of nonzero normal vectors. It is easy to see that for the induced orientation of the border set ∂D in (6) that field consists of *outer normals.*

In the plane $\mathbb{R}^2$ the analogue of formula (6) is of the form

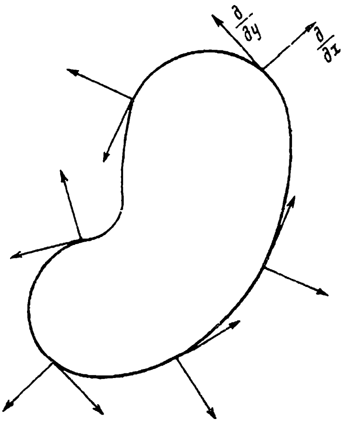
$$\iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy = \int_{\partial D} P\,dx + Q\,dy \tag{7}$$

and is called the **Green's formula.** Here the orientation of the border set ∂D, i.e. the direction of tracing the boundary is given by the requirement that the domain D should remain on the left.

Finally, on the straight line $\mathbb{R}$ formula (3) goes into the **Newton-Leibniz formula**

$$\int_a^b f'(x)\,dx = f(b) - f(a). \tag{8}$$

(Recall that the integral over an oriented zero-dimensional manifold is given by formula (3) of Lecture 25 and that by



definition the point b enters the border set of the closed interval $[a, b]$ with a plus sign, and the point a does with a minus sign.)

The notion of a domain with regular boundary, together with Theorem 1, allows an important generalization.

Let $\mathbb{R}^n_{(-)}$ be a half-space of $\mathbb{R}^n$ consisting of points $\mathbf{x} = (x^1, \ldots, x^n)$ for which $x^1 \leqslant 0$, and let $\mathbb{R}^{n-1}_0$ be its *border set* consisting of the points $\mathbf{x}$ for which $x^1 = 0$.

A subset $U \subset \mathbb{R}^n$ is said to be *∂-open* if it is an open set in either $\mathbb{R}^n$ or $\mathbb{R}^n_{(-)}$, i.e. if there is an open set U' in $\mathbb{R}^n$ such that either $U = U'$ or $U = U' \cap \mathbb{R}^n_{(-)}$. A mapping $\varphi \colon U \to \mathbb{R}^m$ defined on a ∂-open set $U \subset \mathbb{R}^n$ is said to be *smooth* if there is an open set $U' \subset \mathbb{R}^n$ and a smooth mapping $\varphi' \colon U' \to \mathbb{R}^m$ such that $U \subset U'$ and $\varphi' \mid_U = \varphi$. This is equivalent to saying that functions giving a mapping φ have continuous partial derivatives of all the necessary orders (subject to the condition, in the case where $U \subset \mathbb{R}^n_{(-)}$ and $U \cap \mathbb{R}^{n-1}_0 \neq \varnothing$, that at the points of $U \cap \mathbb{R}^{n-1}_0$ differentiation with respect to x^1 is understood as differentiation from the right). A mapping $\varphi \colon U \to V$ of ∂-open sets is said to be a *diffeomorphism* if it is bijective smooth and the inverse mapping $\varphi^{-1} \colon V \to U$ is also smooth.

Let $\mathscr{X}$ be an arbitrary set. A pair (U, h) consisting of a subset $U \subset \mathscr{X}$ and a bijective mapping $h \colon U \to h(U)$ on a ∂-open set $h(U) \subset \mathbb{R}^n$ is called a *∂-chart* in $\mathscr{X}$. Two ∂-charts (U, h) and (V, k) are said to be *compatible* if either $U \cap V = \varnothing$ or $U \cap V \neq \varnothing$ and

(a) both sets $h(U \cap V)$ and $k(U \cap V)$ are ∂-open subsets of $\mathbb{R}^n$;

(b) the mapping

$$k \circ h^{-1} \colon h(U \cap V) \to k(U \cap V)$$

is a diffeomorphism. It is clear that every chart in $\mathscr{X}$ (in the sense of Definition 1 of Lecture 6) is a ∂-chart and compatible charts (in the sense of Definition 2 of Lecture 6) are compatible as ∂-charts.

By an *atlas of ∂-charts* on a set $\mathscr{X}$ is meant (cf. Definition 3 of Lecture 6) a set of pairwise compatible ∂-charts (U_α, h_α) with the property that their *supports* U_α cover $\mathscr{X}$. It is easy to see (cf. Proposition 1 of Lecture 6 and its Corollary 1) that any atlas of ∂-charts A is contained in a unique maximal atlas which consists of all ∂-charts compatible with ∂-charts of A.

Definition 2. A set $\mathscr{D}$ in which a maximal atlas of ∂-charts is given is called a *smooth ∂-manifold* (and the charts of the atlas are *smooth charts*).

Of course, every manifold $\mathscr{X}$ is a ∂-manifold.

Note that just as in Lecture 6 we assume the dimension n of $\mathbb{R}^n$ to be fixed (it is termed the *dimension* of a ∂-manifold $\mathscr{D}$ and designated $\dim \mathscr{D}$) as well as the smoothness class C^r of all the mappings under consideration, where r is either a nonnegative integer or ∞ or ω. (This class is called the *smoothness class* of a ∂-manifold $\mathscr{D}$; as a rule, we shall assume that $r = \infty$.)

The topology is introduced into ∂-manifolds the way it is introduced into manifolds, i.e. a subset O in a ∂-manifold $\mathscr{D}$ is considered to be open if and only if for any smooth chart (U, h) the set $h(O \cap U)$ is ∂-open in $\mathbb{R}^n$. Thus every ∂-manifold turns out to be a topological space (satisfying the first countability axiom and, assuming its being Hausdorff, to be locally compact).

Definition 3. By an *interior chart* of a ∂-manifold $\mathscr{D}$ is meant a ∂-chart (U, h) for which the set $h(U)$ is open in $\mathbb{R}^n$. A point $p \in \mathscr{D}$ is said to be an *interior point* of a ∂-manifold $\mathscr{D}$ if there is an interior chart (U, h) in $\mathscr{D}$ such that $p \in U$. The set of all interior points of a ∂-manifold $\mathscr{D}$ is designated $\operatorname{int} \mathscr{D}$ (note that the first letter is small!) and is called the *interior* of $\mathscr{D}$.

Clearly the set $\operatorname{int} \mathscr{D}$ is open in $\mathscr{D}$ and is a smooth manifold (with an atlas consisting of all interior charts).

Problem 1. Prove that $\operatorname{int} \mathscr{D}$ is not empty (if $\mathscr{D}$ is not empty) and, moreover, that

$$\overline{\operatorname{int} \mathscr{D}} = \mathscr{D}.$$

Definition 4. By a *boundary chart* of a ∂-manifold $\mathscr{D}$ is meant its ∂-chart (U, h) such that $h(U) \subset \mathbb{R}^n_{(-)}$ and $h(U) \cap \mathbb{R}^{n-1}_0 \neq \varnothing$. A point $p \in \mathscr{D}$ is said to be a *point of a border set* of $\mathscr{D}$ if there is a boundary chart (U, h) in $\mathscr{D}$ such that $p \in U$ and $h(p) \in h(U) \cap \mathbb{R}^{n-1}_0$. The set (possibly empty) of all points of the border set is designated $\partial\mathscr{D}$ and called the *border set* of the ∂-manifold $\mathscr{D}$. By definition

$$\mathscr{D} = \partial\mathscr{D} \cup \operatorname{int} \mathscr{D}.$$

Proposition 1. *No interior point is a point of a border set, and conversely no point of a border set is an interior point:*

$$\partial\mathscr{D} \cap \operatorname{int}\mathscr{D} = \varnothing.$$

Proof. The statement that $p \in \partial\mathscr{D} \cap \operatorname{int}\mathscr{D}$ implies that there is an open set U in $\mathscr{D}$ that contains the point p and mappings $h\colon U \to \mathbb{R}^n_{(-)}$ and $k\colon U \to \mathbb{R}^n$ such that the pair (U, h) is a boundary chart and the pair (U, k) an interior chart of $\mathscr{D}$. Consequently $k(U)$ is an open set in $\mathbb{R}^n$ and $h(U)$ is not. Since the mapping $h \circ k^{-1}$ is a diffeomorphism of $k(U)$ onto $h(U)$, this contradicts Lemma 1 to be proved below. Therefore the point p cannot exist, and hence $\partial\mathscr{D} \cap \operatorname{int}\mathscr{D} = \varnothing$. □

Lemma 1. *Let $\varphi\colon O \to \mathbb{R}^n$ be a smooth C^r-mapping $(r \geqslant 1)$ of an open set $O \subset \mathbb{R}^n$ into $\mathbb{R}^n$. If at every point $\mathbf{x} \in O$ the Jacobian D_φ of φ is nonzero, then $\varphi(O)$ is an open set.*

Proof. By the inverse mapping theorem (see Lecture 6) a point $\mathbf{x} \in O$ has a neighbourhood $U \subset O$ which is mapped onto some neighbourhood V of a point $\varphi(\mathbf{x})$. Since $V \subset \varphi(O)$ it follows that $\varphi(\mathbf{x}) \in \operatorname{Int}\varphi(O)$ and since it is true for any point $\mathbf{x} \in O$, we have $\varphi(O) = \operatorname{Int}\varphi(O)$, i.e. $\varphi(O)$ is an open set. □

Proposition 1 is true for $r = 0$ as well (for topological manifolds). The corresponding analogue of Lemma 1 (known as the **Brouwer theorem on the domain invariance**) states that if a continuous mapping $\varphi\colon O \to \mathbb{R}^n$ of an open set $O \subset \mathbb{R}^n$ into $\mathbb{R}^n$ is a moneomorphism (a homeomorphism onto $\varphi(O)$), then the set $\varphi(O)$ is open. Unfortunately, there is no room to present here a rather lengthy and tedious proof of this statement.

Corollary 1. *A border set $\partial\mathscr{D}$ of an arbitrary ∂-manifold $\mathscr{D}$ is closed in $\mathscr{D}$.*

Proof. According to Proposition 1 $\partial\mathscr{D} = \mathscr{D} \setminus \operatorname{int}\mathscr{D}$ and $\operatorname{int}\mathscr{D}$ is open in $\mathscr{D}$. □

Corollary 2. *The equation $\partial\mathscr{D} = \varnothing$ holds if and only if the ∂-manifold $\mathscr{D}$ is a manifold.*

Proof. Clearly $\mathscr{D}$ is a manifold if and only if $\operatorname{int}\mathscr{D} = \mathscr{D}$. □

On the basis of Corollary 2 smooth manifolds are also called *manifolds without border sets*. Accordingly, ∂-manifolds with $\partial\mathscr{D} \neq \varnothing$ are called *manifolds with border sets*. However, the term "manifold with border set" is often used as a synonym of the term "∂-manifold". (But where this terminological licence may lead to confusion one speaks of "manifold with a nonempty border set" or of "manifold with or without border set".)

Of particular importance are compact manifolds without border set. Such manifolds are called *closed*.

For every boundary chart $(U, h) = (U, x^1, \ldots, x^n)$ of a ∂-manifold $\mathscr{D}$ the pair $(U \cap \partial\mathscr{D}, h_{U\cap\partial\mathscr{D}})$ is by virtue of the identification $\mathbb{R}_0^{n-1} = \mathbb{R}^{n-1}$ a chart in $\partial\mathscr{D}$ (with local coordinates $x^2, \ldots, x^n$), and these charts make up an atlas on $\partial\mathscr{D}$. This shows that the *border set $\partial\mathscr{D}$ of an arbitrary n-dimensional ∂-manifold $\mathscr{D}$ is an $(n - 1)$-dimensional manifold without border set.*

It is clear that *any domain D with regular boundary is a ∂-manifold with border set ∂D*. (For this ∂-manifold ∂-charts are the charts of an encompassing manifold $\mathscr{X}$ contained in Int D or intersections of charts adapted to D with D.) The dimension of this ∂-manifold is equal to the dimension n of the manifold $\mathscr{X}$.

A mapping $f\colon \mathscr{D} \to \mathscr{X}$ of a ∂-manifold $\mathscr{D}$ into a ∂-manifold $\mathscr{X}$ (we shall need only the case where $\mathscr{X}$ has no border set) is said to be *smooth* if it is continuous and for any two ∂-charts (U, h) in $\mathscr{D}$ and (V, k) in $\mathscr{X}$ for which $fU \subset V$ the composite mapping

$$k \circ f \circ h^{-1}\colon h(U) \to k(V)$$

is smooth.

Let a manifold $\mathscr{X}$ have no border set (in the case where $\partial\mathscr{X} \neq \varnothing$ some complications arise we do not wish to go into at this point), and let $\mathscr{D} \subset \mathscr{X}$. If the embedding $i\colon \mathscr{D} \to \mathscr{X}$ is smooth and the manifolds int $\mathscr{D}$ and $\partial\mathscr{D}$ are embedded submanifolds of $\mathscr{X}$, then $\mathscr{D}$ is said to be an (*embedded*) *∂-submanifold* of $\mathscr{X}$.

For example, any domain with regular boundary is a ∂-submanifold.

Problem 2. Prove that if the dimension of a ∂-submanifold $\mathscr{D}$ of a manifold $\mathscr{X}$ is equal to the dimension of $\mathscr{X}$, then $\mathscr{D}$ is a domain with regular boundary in $\mathscr{X}$.

By a *differential form* ω *of degree* m on a ∂-manifold $\mathscr{D}$ we shall mean a form on int $\mathscr{D}$ such that for any boundary chart (U, h) the coefficients $\omega_{i_1 \dots i_m}$ of the form in a chart $(U \cap \operatorname{int} \mathscr{D},\ h|_{U \cap \operatorname{int} \mathscr{D}})$ of the manifold int $\mathscr{D}$ are restrictions of some (obviously uniquely defined) smooth functions given on U. These functions will be denoted by the same symbols $\omega_{i_1 \dots i_m}$ and called the *coefficients of a form* ω *in a chart* (U, h). Restrictions of the coefficients $\omega_{i_1 \dots i_m}$, $i_1, \dots, i_m \neq 1$, to $U \cap \partial\mathscr{D}$ are obviously the coefficients of some form of degree $m - 1$ on $\partial\mathscr{D}$ which will be designated $\omega|_{\partial\mathscr{D}}$.

In the case where $\mathscr{D}$ is a ∂-submanifold of a manifold $\mathscr{X}$ without border set (for example, a domain with regular boundary), for any form ω on $\mathscr{X}$ the form $j^*\omega$, where j: int $\mathscr{D} \to \mathscr{X}$ is an embedding, is obviously a form on $\mathscr{D}$. For this form $j^*\omega|_{\partial\mathscr{D}} = i^*\omega$, where i is an embedding $\partial\mathscr{D} \to \mathscr{X}$.

Problem 3. Prove that any form on $\mathscr{D}$ is $j^*\omega$, where ω is a form on $\mathscr{X}$.

A manifold with border set $\mathscr{D}$ is said to be *orientable* (*oriented*) if so is the manifold int $\mathscr{D}$. On an oriented ∂-manifold $\mathscr{D}$ a boundary chart (U, h) is said to be *positive* if so is the interior chart $(U \cap \operatorname{int} \mathscr{D},\ h|_{U \cap \operatorname{int} \mathscr{D}})$.

A differential form on a ∂-manifold $\mathscr{D}$ is said to be *finite* if it is zero outside some compact set $C \subset \mathscr{D}$. If a Hausdorff and paracompact ∂-manifold $\mathscr{D}$ is oriented, then for any finite form ω of degree $n = \dim \mathscr{D}$ an integral

$$\int_{\operatorname{int} \mathscr{D}} \omega$$

is defined. We shall call it the *integral of* ω *over* $\mathscr{D}$ and designate

$$\int_{\mathscr{D}} \omega. \tag{9}$$

If $\mathscr{D}$ is a domain D with regular boundary in a manifold $\mathscr{X}$ and the form ω is a restriction of some (finite if the domain D is not compact) form ω' to $\mathscr{X}$, then since the border set ∂D is a null set by virtue of the Sard theorem, integral (9) is equal to the integral

$$\int_D \omega'$$

of the form ω' over D.

The construction of an induced orientation of a border set for domains with regular boundary can be carried over word for word to any oriented ∂-manifolds. In what follows when speaking about the border set of an oriented ∂-manifold $\mathscr{D}$ we shall always assume that it is provided with an induced orientation.

In particular, for any finite form ω of degree $n-1$ on an oriented n-dimensional Hausdorff and paracompact ∂-manifold $\mathscr{D}$ this makes it possible to speak about the integral

$$\int_{\partial\mathscr{D}} \omega \,|_{\partial\mathscr{D}}$$

which for brevity we shall simply denote by

$$\int_{\partial\mathscr{D}} \omega. \tag{10}$$

Theorem 1′ (Stokes' theorem for manifolds with border set). *For any finite form ω of degree $n-1$ on an n-dimensional Hausdorff paracompact and oriented ∂-manifold $\mathscr{D}$ there is*

$$\int_{\mathscr{D}} d\omega = \int_{\partial\mathscr{D}} \omega. \tag{11}$$

The **proof** of Theorem 1′ repeats virtually word for word that of Theorem 1 and it will be left to the reader. □

For $\partial\mathscr{D} = \varnothing$ it is appropriate to assume integral (10) to be zero. By virtue of this convention Corollary 1 to Theorem 1 turns out to be a special case of Theorem 1′.

In the case where $\mathscr{D}$ is a ∂-submanifold of a manifold $\mathscr{X}$, for any form ω of degree n on $\mathscr{X}$ (notice that here n

is the dimension of $\mathscr{D}$ rather than $\mathscr{X}$!) with the property that the form $j^*\omega$ on $\mathscr{D}$ is finite (where as before j is an embedding $\mathscr{D} \to \mathscr{X}$) the integral

$$\int_{\mathscr{D}} j^*\omega$$

is denoted by

$$\int_{\mathscr{D}} \omega \tag{12}$$

(and forms ω on $\mathscr{X}$ for which the form $j^*\omega$ is finite are called forms *finite on* $\mathscr{D}$).

By virtue of this convention the following theorem, a generalization of Theorem 1, is obviously true.

Theorem 1″ (Stokes' theorem for ∂-submanifolds). *For any n-dimensional oriented ∂-submanifold $\mathscr{D}$ of a Hausdorff and paracompact manifold $\mathscr{X}$ and any form ω finite on $\mathscr{D}$ of degree $n-1$ on $\mathscr{X}$ formula* (11) *holds.* □

Two-dimensional ∂-submanifolds of $\mathbb{R}^3$ are called *surfaces with border set*. For any such surface $\mathscr{D}$ and any form

$$\omega = P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy$$

in $\mathbb{R}^3$ integral (12) is denoted by

$$\iint_D P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy \tag{13}$$

(cf. (5)). In this case formula (11) (for forms $\omega = P\,dx + Q\,dy + R\,dz$) becomes therefore

$$\iint_{\mathscr{D}} \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy\,dz + \left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z}\right) dz\,dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\,dy = \int_{\partial\mathscr{D}} P\,dx + Q\,dy + R\,dz. \tag{14}$$

Formula (14) is known as **Stokes' formula for surface integrals in $\mathbb{R}^3$.**

Remark 3. It is interesting that the name "Stokes' formula" first applied to formula (14) and then carried over to its generalizations and analogues (3) and (11) originated as a result of misunderstanding. Formula (14) came to be known at Cambridge University, England, in the mid 19th century and on the suggestion of the well-known physicist and mathematician Thompson (who may have been the first to devise it) was included in question cards in mathematics for students of the University. It was ascribed to Stokes, first only by students, just because Stokes was at the time the chairman of the examining board who signed the question cards.

Theorems 1′ and 1″ may be combined in one general theorem.

Let $\mathscr{D}$ and $\mathscr{X}$ be arbitrary ∂-manifolds, and let $\gamma\colon \mathscr{D} \to \mathscr{X}$ be a smooth mapping carrying int $\mathscr{D}$ into int $\mathscr{X}$. Then for any form ω on $\mathscr{X}$ the form $\gamma^*\omega$ (defined on int $\mathscr{D}$) is easily seen to be a form on $\mathscr{D}$. If therefore it is finite and of degree $n = \dim \mathscr{D}$ and the manifold $\mathscr{D}$ is orientable, then an integral

$$\int_{\mathscr{D}} \gamma^*\omega \tag{15}$$

is defined.

Smooth mappings $\gamma\colon \mathscr{D} \to \mathscr{X}$ carrying int $\mathscr{D}$ into int $\mathscr{X}$ for which the ∂-manifold $\mathscr{D}$ is oriented and n-dimensional are called *n-dimensional singular submanifolds* of the ∂-manifold $\mathscr{X}$, while forms ω on $\mathscr{X}$ for which the form $\gamma^*\omega$ is finite are called forms *finite on* γ and integral (15) is the *integral of* ω *over* γ. Accordingly we designate integral (15) as

$$\int_{\gamma} \omega. \tag{16}$$

A "mapping" $\varnothing \to \mathscr{X}$ of an empty set $\varnothing$ into $\mathscr{X}$ is also considered to be an n-dimensional singular submanifold. Integral (16) over such a submanifold is assumed to be zero.

By virtue of this convention a singular submanifold

$$\gamma|_{\partial\mathscr{D}}\colon \partial\mathscr{D} \to \mathscr{X}$$

is defined for any singular submanifold $\gamma\colon \mathscr{D} \to \mathscr{X}$. This singular submanifold is called the *border set* of γ and designated $\partial\gamma$.

Theorem 1‴ (Stokes' theorem for singular submanifolds). *For any n-dimensional singular submanifold γ of a manifold $\mathscr{X}$ and any form ω finite on γ of degree $n-1$ on $\mathscr{X}$ a formula*

$$\int_{\gamma} d\omega = \int_{\gamma} \omega \tag{17}$$

holds.

Proof. It suffices to apply Theorem 1′ to the form $\gamma^*\omega$ on $\mathscr{D}$. ☐

Theorem 1‴ reduces to Theorem 1′ when γ is an identity mapping id: $\mathscr{D} \to \mathscr{D}$ and to Theorem 1″ when γ is an embedding $\mathscr{D} \to \mathscr{X}$.

Integrals over one-dimensional singular submanifolds of $\mathbb{R}^n$, $n \geqslant 2$, allow quite an elementary treatment. Let for definiteness $n = 2$, and let $\mathscr{D}$ be a closed interval $[a, b]$ and hence a singular mapping $\gamma\colon \mathscr{D} \to \mathbb{R}^2$ be nothing but a plane smooth curve

$$x = x(t),\ y = y(t),\ a \leqslant t \leqslant b \tag{18}$$

(subject in general to no regularity requirements). A trivial interpretation of the definitions shows that for any form $P\,dx + Q\,dy$ on $\mathbb{R}^2$ the integral over curve (18) is given by the formula

$$\int_{\gamma} P\,dx + Q\,dy = \int_a^b [P(x(t),\ y(t))\,x'(t) + Q(x(t),\ y(t))\,y'(t)]\,dt \tag{19}$$

which may be taken to be its definition. Integrals of such a type occur frequently in calculus problems and are called *line integrals of the second kind.*

For $n = 3$ they are of the form

$$\int_\gamma P\,dx + Q\,dy + R\,dz = \int_a^b [P(x(t), y(t), z(t))\,x'(t) + Q(x(t), y(t), z(t))\,y'(t) + R(x(t), y(t), z(t))\,z'(t)]\,dt$$

and similarly for any n.

[Line (or curvilinear) integrals of the first kind were introduced in the preceding lecture. They are integrals of densities rather than forms.]

Notice that integral (19) makes sense not only for smooth but also for **piecewise smooth** curves γ.

In the case where $\mathscr{D}$ is an oriented circle $\mathbb{S}^1$ the singular manifold γ is called a *closed curve*. Choosing a closed interval $[a, b]$, a point $s_0 \in \mathbb{S}^1$ and a mapping $\alpha\colon [a, b] \to \mathbb{S}^1$ carrying the points a and b into the point s_0 and diffeomorphically mapping (while preserving orientation) an open interval (a, b) onto an arc $\mathbb{S}^1 \setminus \{s_0\}$ we can identify every such curve with a curve $\beta = \gamma \circ \alpha\colon [a, b] \to \mathbb{R}^n$ having the property that $\beta(a) = \beta(b)$. (On this basis curves $\beta\colon [a, b] \to \mathbb{R}^n$ for which $\beta(a) = \beta(b)$ are also called *closed curves*, cf. Lecture 1.)

This identification is compatible with integration, i.e. for any closed curve γ an integral over γ (generally designated $\oint\limits_\gamma$ to stress that γ is a closed curve) is equal to a line integral over β.

When $\mathscr{D}$ consists of several oriented circles the singular manifold γ is a system $\gamma_1, \ldots, \gamma_n$ of closed curves, and a formula

$$\oint_\gamma = \oint_{\gamma_1} + \ldots + \oint_{\gamma_n}$$

holds for the corresponding integral. In particular, for any compact plane domain $\mathscr{D} \subset \mathbb{R}^2$ with regular (or *piecewise regular* in the obvious sense) boundary and any form $P\,dx + Q\,dy$ on $\mathbb{R}^2$ this identifies the integral

$\oint\limits_{\partial\mathscr{D}} P\,dx + Q\,dy$ of the Green's formula (7) with the line integral $\oint\limits_{\partial\mathscr{D}} P\,dx + Q\,dy$.

Thus both integrals in the Green's formula (7) allow quite an elementary treatment, which cannot be said about, for example, the integral on the right of the Gauss-Ostrogradskii formula (6) or the one on the left of Stokes' formula (14).

Lecture 28

Operators of vector analysis · Consequences of the identity $d \circ d = 0$ *· Consequences of differentiation formulas for products · The Laplacian and the Beltrami operator · The flux of a vector field · The Gauss-Ostrogradskii formulas for divergence and Green's formulas · Convergence as the density of sources · The Stokes formula for circulation · The Gauss-Ostrogradskii formula for rotation · The generalized Gauss-Ostrogradskii formula*

In this lecture we shall study the case of forms in the space $\mathbb{R}^3$ which is especially important for applications.

The availability in $\mathbb{R}^3$ of a fixed coordinate system and of a standard Euclidean metric allows us to identify linear differential forms $P\,dx + Q\,dy + R\,dz$ on $\mathbb{R}^3$ (as well as the second-degree forms $P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy$) with vector fields

$$(1) \qquad \mathbf{u} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k},$$

where $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are vectors of the standard basis in $\mathbb{R}^3$. This leads to a very rich theory known as **vector analysis** (or **field theory**). Although vector analysis is of minimum theoretical importance, we shall nevertheless develop it quite explicity since it plays an important role in physical and technical applications connected with hydrodynamics and electromagnetism (suffice it to say that the Maxwell equations for an electromagnetic field can be written most elegantly, unless we use the four-dimensional formalism of special relativity theory, using differential operators of vector analysis).

The principal field of vector analysis is the algebra of smooth functions F defined on some open set $W \subset \mathbb{R}^3$ (which is considered to be fixed throughout what follows) and the F-module $\mathfrak{a}$ of vector fields on W.

The identification of the module $\mathfrak{a}$ with the modules Ω^1 and Ω^2 of differential forms of degrees 1 and 2 (and the identification of the algebra F considered as a module over itself with the module Ω^3 of forms $f\,dx \wedge dy \wedge dz$ of degree 3) allows the operators of exterior differentiation

$$\mathsf{F} = \Omega^0 \xrightarrow{d} \Omega^1 \xrightarrow{d} \Omega^2 \xrightarrow{d} \Omega^3$$

to be interpreted as the operators

$$\mathsf{F} \to \mathfrak{a} \to \mathfrak{a} \to \mathsf{F}.$$

The first of these operators (the one from F to $\mathfrak{a}$) is designated grad. It associates an arbitrary smooth function F with a vector field

$$\operatorname{grad} F = \frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j} + \frac{\partial F}{\partial z}\mathbf{k} \tag{2}$$

called the *gradient* of F.

The second operator (the one from $\mathfrak{a}$ to $\mathfrak{a}$) is designated curl. It associates the vector field (1) with a field

$$\operatorname{curl} \mathbf{u} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\mathbf{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\mathbf{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\mathbf{k} \tag{3}$$

called the *curl* (or *rotation*) of a field $\mathbf{u}$.

Formula (3) can be written in the following mnemonic form

$$\operatorname{curl} \mathbf{u} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}. \tag{4}$$

The third operator (the one from $\mathfrak{a}$ to F) is designated div. It associates the vector field (1) with a function

$$\operatorname{div} \mathbf{u} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \tag{5}$$

called the *divergence* of $\mathbf{u}$.

Besides these operators (and the operation of multiplication by a function) also defined on the module $\mathfrak{a}$ are operations of scalar and vector multiplication which associate vector fields

$$\mathbf{u} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k} \text{ and } \mathbf{v} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}$$

with a function

$$\mathbf{uv} = PX + QY + RZ$$

and a vector field

$$\mathbf{u} \times \mathbf{v} = (QZ - RY)\,\mathbf{i} + (RX - PZ)\,\mathbf{j} + (PY - QX)\,\mathbf{k} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ P & Q & R \\ X & Y & Z \end{vmatrix}$$

respectively.

All these operators (and operations) are connected with one another by numerous identities.

In the first place, the identity $d \circ d = 0$ gives us two identities

(6) $\quad$ curl grad $F = 0$ and div curl $\mathbf{u} = 0$

which are easy to verify by a straightforward computation.

Fields of the form grad F are called *potential fields* and those of the form curl $\mathbf{u}$ are *solenoidal fields*. If $\mathbf{u}$ = grad F, then F is said to be the *potential* of the field $\mathbf{u}$ and if $\mathbf{v}$ = curl $\mathbf{u}$, then the field $\mathbf{u}$ is said to be a *vector potential* of the field $\mathbf{v}$. If W is connected, then the potential F is defined by the field $\mathbf{u}$ up to a constant term.

A field $\mathbf{u}$ is said to be *irrotational* if curl $\mathbf{u} = 0$ and to be a *source-free field* if div $\mathbf{u} = 0$. The vector potential $\mathbf{u}$ of the field curl $\mathbf{u}$ is defined up to an irrotational field.

By identity (6) *any potential field is irrotational and any solenoidal field is a source-free field.* [The quotient space of a vector space of irrotational fields with respect to a subspace of potential fields is nothing but a one-dimensional de Rham cohomology group H^1W of a domain W and the quotient space of source-free fields with respect to a subspace of solenoidal fields is a group H^2W. Therefore all irrotational fields are potential fields if and only if $H^1W = 0$, and all source-free fields are solenoidal if and only if $H^2W = 0$.]

Example 1. A field of the form

$$\mathbf{u} = f(r)\,\mathbf{r}, \quad \mathbf{r} \neq 0,$$

where $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and $r = |\mathbf{r}| = \sqrt{x^2 + y^2 + z^2}$, is called a *central (force) field.* For this field

$$P = f(r)\,x, \quad Q = f(r)\,y, \quad R = f(r)\,z$$

(and $W = \mathbb{R}^3 \setminus \{0\}$). Since

$$\frac{\partial r}{\partial x} = \frac{x}{r}, \quad \frac{\partial r}{\partial y} = \frac{y}{r}, \quad \frac{\partial r}{\partial z} = \frac{z}{r},$$

the matrix

$$\left\| \begin{matrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} & \frac{\partial P}{\partial z} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} & \frac{\partial Q}{\partial z} \\ \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} & \frac{\partial R}{\partial z} \end{matrix} \right\|$$

is of the form

$$(7) \quad \left\| \begin{matrix} f'(r)\frac{x^2}{r} + f(r) & f'(r)\frac{xy}{r} & f'(r)\frac{xz}{r} \\ f'(r)\frac{xy}{r} & f'(r)\frac{y^2}{r} + f(r) & f'(r)\frac{yz}{r} \\ f'(r)\frac{xz}{r} & f'(r)\frac{yz}{r} & f'(r)\frac{z^2}{r} + f(r) \end{matrix} \right\|.$$

The symmetry of the matrix implies that curl $\mathbf{u} = 0$, i.e. that *every central field is an irrotational field.* Moreover, since $H^1(\mathbb{R}^3 \setminus \{0\}) = H^1(S^2) = 0$, *every central field is a potential field.* (The corresponding potential F is defined as follows:

$$F(\mathbf{r}) = \int_1^r r f(r)\,dr,$$

where the lower limit of integration may of course be any other positive number.)

If, in particular, $f(r) = -\frac{m_0}{r^3}$ and hence $|\mathbf{u}| = \frac{m_0}{r^2}$ **(the gravitational field of a point of mass** m_0), then $F(\mathbf{r}) = \frac{m_0}{r}$ (up to a constant). This potential is termed a *Newtonian potential.*

The divergence of a central field is the sum

$$f'(r)\frac{x^2+y^2+z^2}{r}+3f(r)=rf'(r)+3f(r)$$

of the diagonal elements of matrix (7). Solving the differential equation $rf'(r) + 3f(r) = 0$ we immediately get $f(r) = \frac{C}{r^3}$. Thus *a central field is a gravitational field (a field of Newtonian potential) if and only if it has no sources* (in the domain $\mathbb{R}^3 \setminus \{0\}$, where it is defined).

By virtue of our identifications the outer product $\omega \wedge \theta$ of differential forms goes into a vector product of fields for $\deg \omega = 1$ and $\deg \theta = 1$ and into their scalar product for $\deg \omega = 1$ and $\deg \theta = 2$ (for $\deg \omega = 0$ it is a product of the field θ by the function ω). Therefore the formula

$$d(f\omega) = f\,d\omega + df \wedge \omega$$

yields relations

(8) $$\operatorname{grad}(FG) = F \operatorname{grad} G + G \operatorname{grad} F,$$

(9) $$\operatorname{curl}(F\mathbf{u}) = F \operatorname{curl} \mathbf{u} + \operatorname{grad} F \times \mathbf{u},$$

(10) $$\operatorname{div}(F\mathbf{u}) = F \operatorname{div} \mathbf{u} + \operatorname{grad} F \cdot \mathbf{u},$$

and the formula

$$d(\omega \wedge \theta) = d\omega \wedge \theta - \omega \wedge d\omega, \quad \omega, \theta \in \Omega^1,$$

yields

(11) $$\operatorname{div}(\mathbf{u} \times \mathbf{v}) = (\operatorname{curl} \mathbf{u})\,\mathbf{v} - \mathbf{u} \operatorname{curl} \mathbf{v}.$$

Of course, formulas (8) to (11) can easily be obtained by a straightforward computation.

The mnemonic formula (4) suggests that a symbolic vector field

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}$$

should be introduced into consideration. Then the field curl **u** is a vector product $\nabla \times \mathbf{u}$ of a field ∇ by a field **u**:

$$\operatorname{curl} \mathbf{u} = \nabla \times \mathbf{u},$$

the function div **u** is a scalar product $\nabla\mathbf{u}$ of the field ∇ by the field **u**:

$$\operatorname{div} \mathbf{u} = \nabla\mathbf{u},$$

and the field grad F, with the scalar factor written on the right, is a product ∇F of the function F by the field ∇:

$$\operatorname{grad} F = \nabla F.$$

These representations of the operators curl, div, and grad by a symbolic operator ∇ allow relations (8)-(11) to be written as one identity

$$(12) \qquad \nabla \textcircled{1} (\alpha \textcircled{1} \gamma) = \nabla \textcircled{1} (\overset{\downarrow}{\alpha} \textcircled{2} \beta) + \nabla \textcircled{1} (\alpha \textcircled{2} \overset{\downarrow}{\beta}),$$

where α and β are either functions or fields, ① and ② are two of the three possible multiplications (multiplication by a function, scalar or vector multiplication) and $\downarrow$ marks the factor subjected to the action of the operator ∇. Indeed, when $\alpha = F$ and $\beta = G$ this is obviously formula (8), and when $\alpha = F$ and $\beta = \mathbf{u}$ we have, depending on the choice of multiplications, formulas (9) and (10). When $\alpha = \mathbf{u}$ and $\beta = \mathbf{v}$, on the assumption that ① is a scalar multiplication and ② is a vector one, formula (12) is of the form

$$\nabla (\mathbf{u} \times \mathbf{v}) = \nabla (\overset{\downarrow}{\mathbf{u}} \times \mathbf{v}) + \nabla (\mathbf{u} \times \overset{\downarrow}{\mathbf{v}}).$$

But by virtue of the skew commutativity of a triple scalar product (and the commutativity of a scalar product)

$$\begin{aligned} \nabla (\overset{\downarrow}{\mathbf{u}} \times \mathbf{v}) &= \nabla\overset{\downarrow}{\mathbf{u}}\mathbf{v} = \mathbf{v}\nabla\overset{\downarrow}{\mathbf{u}} = \mathbf{v}\nabla\mathbf{u} = \mathbf{u}(\nabla \times \mathbf{u}) \\ &= (\nabla \times \mathbf{u})\,\mathbf{v} \end{aligned}$$

and

$$\nabla (\mathbf{u} \times \overset{\downarrow}{\mathbf{v}}) = \nabla\mathbf{u}\overset{\downarrow}{\mathbf{v}} = -\mathbf{u}\nabla\overset{\downarrow}{\mathbf{v}} = -\mathbf{u}\nabla\mathbf{v} = -\mathbf{u}(\nabla \times \mathbf{v}).$$

In this case formula (12) reduces therefore to formula (11).

Notice that when $\alpha = \mathbf{u}$ and $\beta = \mathbf{v}$ two more cases remain unconsidered in (12):

$$\nabla(\mathbf{uv}) = \nabla(\overset{\downarrow}{\mathbf{u}}\mathbf{v}) + \nabla(\mathbf{u}\overset{\downarrow}{\mathbf{v}}) \tag{13}$$

and

$$\nabla \times (\mathbf{u} \times \mathbf{v}) = \nabla \times (\overset{\downarrow}{\mathbf{u}} \times \mathbf{v}) + \nabla \times (\mathbf{u} \times \overset{\downarrow}{\mathbf{v}}). \tag{14}$$

To decipher these formulas we agree that for any vector field $\mathbf{a} = A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$ we understand by $\mathbf{a}\nabla$ an operator $\alpha \to F$ acting on the field $\mathbf{u} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ by the formula

$$(\mathbf{a}\nabla)\,\mathbf{u} = A\frac{\partial P}{\partial x} + B\frac{\partial Q}{\partial y} + C\frac{\partial R}{\partial z}$$

(thus abandoning the commutativity of the scalar product with respect to the symbolic vector ∇). Multiplying out the vector triple products $\nabla \times (\overset{\downarrow}{\mathbf{u}} \times \mathbf{v})$ and $\nabla \times (\mathbf{u} \times \overset{\downarrow}{\mathbf{v}})$ (by the familiar formula $\mathbf{c} \times (\mathbf{a} \times \mathbf{b}) = (\mathbf{cb})\,\mathbf{a} - (\mathbf{ac})\,\mathbf{b}$; see (19) of Lecture I.22) we then get

$$\nabla \times (\overset{\downarrow}{\mathbf{u}} \times \mathbf{v}) = (\mathbf{v}\nabla)\,\mathbf{u} - (\nabla\mathbf{u})\,\mathbf{v} = (\mathbf{v}\nabla)\,\mathbf{u} - (\operatorname{div}\mathbf{u})\,\mathbf{v}$$

and

$$\nabla \times (\mathbf{u} \times \overset{\downarrow}{\mathbf{v}}) = (\nabla\mathbf{v})\,\mathbf{u} - (\mathbf{u}\nabla)\,\mathbf{v} = (\operatorname{div}\mathbf{v})\,\mathbf{u} - (\mathbf{u}\nabla)\,\mathbf{v}.$$

Since by definition $\nabla \times (\mathbf{u} \times \mathbf{v}) = \operatorname{curl}(\mathbf{u} \times \mathbf{v})$, formula (14) thus becomes

$$\operatorname{curl}(\mathbf{u} \times \mathbf{v}) = (\mathbf{v}\nabla)\,\mathbf{u} - (\mathbf{u}\nabla)\,\mathbf{v} + (\operatorname{div}\mathbf{v})\,\mathbf{u} - (\operatorname{div}\mathbf{u})\,\mathbf{v}. \tag{15}$$

Similarly, by applying to the fields $\mathbf{u} \times \operatorname{curl}\mathbf{v} = \mathbf{u} \times (\nabla \times \mathbf{v})$ and $\mathbf{v} \times \operatorname{curl}\mathbf{u} = \mathbf{v} \times (\nabla \times \mathbf{u})$ the same formula for a vector triple product but rewritten as $\mathbf{c} \times (\mathbf{a} \times \mathbf{b}) = \mathbf{a}\,(\mathbf{cb}) - (\mathbf{ca})\,\mathbf{b}$ we obtain relations

$$\mathbf{u} \times \operatorname{curl}\mathbf{v} = \nabla(\mathbf{u}\overset{\downarrow}{\mathbf{v}}) - (\mathbf{u}\nabla)\,\mathbf{v}$$

and

$$\mathbf{v} \times \operatorname{curl}\mathbf{u} = \nabla(\overset{\downarrow}{\mathbf{u}}\mathbf{v}) - (\mathbf{v}\nabla)\,\mathbf{u}.$$

By virtue of these relations formula (13) becomes

$$(16)\quad \operatorname{grad}(\mathbf{uv}) = \mathbf{u}\times\operatorname{curl}\mathbf{u} + \mathbf{v}\times\operatorname{curl}\mathbf{u} + (\mathbf{v}\nabla)\,\mathbf{u} + (\mathbf{u}\nabla)\,\mathbf{v}.$$

It is remarkable that, as shown by direct computation, formulas (15) and (16) are indeed true (so that (12) is always true when it makes sense).

[The complexity of formulas (15) and (16) is due to the fact that the values of the operator curl on a vector field $\mathbf{u}\times\mathbf{v}$ and of the operator grad on a function $\mathbf{uv}$ cannot be interpreted directly in terms of exterior differentiation of forms.]

Example 2. Let $\mathbf{a}$ be a constant vector. Find a field grad $(\mathbf{ar})$, where, as ever, $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. Clearly, $\operatorname{curl}\cdot\mathbf{a} = 0$ and $(\mathbf{r}\nabla)\,\mathbf{a} = 0$. In addition, as shown by direct computation, $\operatorname{curl}\mathbf{r} = 0$ ($\mathbf{r}$ is a central field) and $(\mathbf{a}\nabla)\,\mathbf{r} = \mathbf{a}$. Hence grad $(\mathbf{ar}) = \mathbf{a}$.

Grouping grad, curl, and div we get

$$\operatorname{curl}\circ\operatorname{grad}\colon \mathsf{F}\to\alpha,\qquad \operatorname{div}\circ\operatorname{grad}\colon \mathsf{F}\to\mathsf{F},$$
$$\operatorname{grad}\circ\operatorname{div}\colon \alpha\to\alpha,\qquad \operatorname{curl}\circ\operatorname{curl}\colon \alpha\to\alpha,$$
$$\operatorname{div}\circ\operatorname{curl}\colon \alpha\to\mathsf{F}.$$

As we know, the operators curl ∘ grad and div ∘ curl are identically zero. [Notice that the operator ∇ reduces this to the statement that a vector product of two identical vectors is zero: $(\operatorname{curl}\circ\operatorname{grad})\,F = \nabla\times\nabla F = (\nabla\times\nabla)\,F = 0$ and $(\operatorname{div}\circ\operatorname{curl})\,\mathbf{u} = \nabla\,(\nabla\times\mathbf{u}) = (\nabla\times\nabla)\,\mathbf{u} = 0.$]

Of paramount interest among the other operators is

$$\operatorname{div}\circ\operatorname{grad} = \nabla^2.$$

This operator is denoted by Δ and called the *Laplace operator* (or *Laplacian*). It transforms every function F into a function

$$\Delta F = \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \frac{\partial^2 F}{\partial z^2}.$$

It is used to write the most important equations of mathematical physics which are treated in a separate course at universities.

It is possible to apply Δ to vector fields as well, applying it to each component individually: if $\mathbf{u} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$, then by definition

$$\Delta\mathbf{u} = (\Delta P)\,\mathbf{i} + (\Delta Q)\,\mathbf{j} + (\Delta R)\,\mathbf{k}.$$

Hence we have

$$\operatorname{curl} \circ \operatorname{curl} = \operatorname{grad} \circ \operatorname{div} - \Delta.$$

Indeed, by the formula for a vector triple product

$$\operatorname{curl}(\operatorname{curl}\mathbf{u}) = \nabla \times (\nabla \times \mathbf{u}) = \nabla(\nabla\mathbf{u}) - (\nabla\nabla)\,\mathbf{u}. \quad \square$$

[This is not a proof, of course, but merely a heuristic confirmation. The real proof, which will be left to the reader, is only by direct computation using formulas (2), (3), and (5).]

It is possible to make up other differential expressions as well. For example, for any two functions F and G we can define a scalar product of their gradients:

$$\operatorname{grad} F \operatorname{grad} G = \frac{\partial F}{\partial x}\frac{\partial G}{\partial x} + \frac{\partial F}{\partial y}\frac{\partial G}{\partial y} + \frac{\partial F}{\partial z}\frac{\partial G}{\partial z}.$$

It is denoted by $\Delta(F, G)$ and called the *Beltrami mixed differential parameter of functions F and G*. In particular, for $F = G$ we obtain a scalar square of the gradient:

$$\Delta(F, F) = (\operatorname{grad} F)^2 = \left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2 + \left(\frac{\partial F}{\partial z}\right)^2.$$

It is denoted by $\Delta_1 F$ and called *Beltrami's first differential parameter of a function F* (cf. Lecture 3).

A scalar triple product of the gradients of three functions is called the *Darboux differential parameter*. This term, however, has almost completely gone out of use now, since this mixed product is nothing else than the Jacobian of the mapping given by the three functions.

Using vector analysis operators the Gauss-Ostrogradskii formula (and the Stokes formula) can be represented in a convenient and compact form.

Let $\mathscr{D}$ be a two-dimensional oriented surface in $\mathbb{R}^3$ (with or without border set). We know (see Remark 2 of Lecture 27) that the orientation of $\mathscr{D}$ is given by some smooth field $\mathbf{n} = \mathbf{n}(M)$ of nonzero normal vectors which, after being normed, may be assumed to consist of unit vectors.

An arbitrary vector field $\mathbf{a}$ defined in an open set W containing a surface $\mathscr{D}$ gives on $\mathscr{D}$ a function a_n which associates every point $M \in \mathscr{D}$ with a projection $a_n(M) = \mathbf{a}(M)\,\mathbf{n}(M)$ of the vector $\mathbf{a}(M)$ onto the direction of $\mathbf{n}(M)$ (the so-called normal component of $\mathbf{a}(M)$).

Definition 1. The integral

$$\int_{\mathscr{D}} a_n\, d\sigma \tag{17}$$

of a function a_n over a surface $\mathscr{D}$ is called the *flux* of a field $\mathbf{a}$ over the surface $\mathscr{D}$. (Here $d\sigma$ is a surface element of $\mathscr{D}$; see Example 2 of Lecture 24.)

Note that the integral of the **first kind** (of density) is meant in formula (17).

If $\mathbf{a}$ is the velocity field of some fluid, then flux (17) is equal to the quantity of fluid passing through the surface $\mathscr{D}$ per unit time.

If $\mathscr{D}$ is an elementary surface, then for any of its parametrizations $\mathbf{r} = \mathbf{r}(u, v)$ *compatible with the orientation*, i.e. such that vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ make up a positive basis for the tangent plane, the normal vector $\mathbf{n}$ is given as follows:

$$\mathbf{n} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{\sqrt{EG - F^2}}$$

(cf. Lecture 4). On the other hand, as we know from Lecture 3, a surface element is given by $d\sigma = \sqrt{EG - F^2}\, du\, dv$. For flux (17) therefore we have

$$\int_{\mathscr{D}} a_n\, d\sigma = \iint_U (\mathbf{a}\mathbf{r}_u\mathbf{r}_v)\, du\, dv = \iint_U \begin{vmatrix} P & Q & R \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} du\, dv, \tag{18}$$

where P, Q, R are the coordinates of the vector $\mathbf{a}$ and U is the domain of $\mathbb{R}^2$ on which a parametrization $\mathbf{r} = \mathbf{r}(u, v)$ is defined. To compute the flux over a nonelementary surface it is necessary to partition it into elementary parts and apply formula (18) to each part.

Now recall that the field $\mathbf{a} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ is associated with a differential form

$$P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy$$

and this form is associated with an integral

$$\int_{\mathscr{Z}} P\,dy \wedge dz + Q\,dz \wedge dx + R\,dx \wedge dy = \iint_{\mathscr{D}} P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy \tag{19}$$

(see formula (13) of Lecture 27). In the case where the surface $\mathscr{D}$ is elementary and parametrized (with parametrization $\mathbf{r} = \mathbf{r}(u, v)$ defined on an open set $U \subset \mathbb{R}^2$), on deciphering the definition of integral (19) we immediately get

$$\iint_U \left[P \begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix} - Q \begin{vmatrix} x_u & z_u \\ x_v & z_v \end{vmatrix} + R \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix} \right] du\,dv = \iint_U \begin{vmatrix} P & Q & R \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} du\,dv.$$

A comparison of this formula with (18) shows that *flux* (17) *of a field* **a** *over a surface* $\mathscr{D}$ *can be expressed by integral* (19):

$$\int_{\mathscr{D}} a_n\,d\sigma = \iint_{\mathscr{D}} P\,dy\,dz + Q\,dz\,dx + R\,dx\,dy. \tag{20}$$

[Proved for elementary surfaces this formula is due to additivity also true for any surfaces.]

Notice that formula (20) expresses the flux in terms of a surface integral of the **second kind.**

When a surface is the border set $\partial\mathscr{D}$ of a domain with regular boundary, integral (19) is nothing but the integral appearing on the right of the Gauss-Ostrogradskii formula (see formula (6) of Lecture 27). Since the integration element on the left is nothing but div **a**, we see that the

Gauss-Ostrogradskii formula may be rewritten as:

$$\iiint\limits_D \operatorname{div} \mathbf{a}\, dV = \oiint\limits_{\partial D} a_n\, d\sigma, \tag{21}$$

where dV denotes $dx\, dy\, dz$ (and the sign $\oiint$ on the right emphasizes that the integration is taken over a closed surface). In words: *the flux of a vector field over the boundary of a domain is equal to the integral of the divergence of the field over the domain.*

In the case where $\mathbf{a} = \operatorname{grad} F$ the normal component a_n is denoted by $\frac{\partial F}{\partial n}$ and called the *directional derivative of F with respect to the normal* (or briefly the *normal derivative*). Since $\operatorname{div} \mathbf{a} = \Delta F$, in this case formula (21) becomes

$$\iiint\limits_D \Delta F\, dV = \oiint\limits_{\partial D} \frac{\partial F}{\partial n}\, d\sigma. \tag{22}$$

When $\mathbf{a} = F \operatorname{grad} G$, according to (10),

$$\begin{aligned} \operatorname{div} \mathbf{a} &= F \operatorname{div} \operatorname{grad} G + \operatorname{grad} F \cdot \operatorname{grad} G \\ &= F \cdot \Delta G + \Delta\,(F,\ G) \end{aligned}$$

and

$$a_n = F \operatorname{grad} G \cdot \mathbf{n} = F\, \frac{\partial G}{\partial n}\,.$$

Therefore

$$\iiint\limits_D [F \cdot \Delta G + \Delta\,(F,\ G)]\, dV = \oiint\limits_{\partial D} F\, \frac{\partial G}{\partial n}\, d\sigma. \tag{23}$$

This formula is called the **first Green's formula**. With $F = G$, it becomes

$$\iiint\limits_D (F \cdot \Delta F + \Delta_1 F)\, dV = \oiint\limits_{\partial D} F\, \frac{\partial F}{\partial n}\, d\sigma$$

and, with $F = 1$, it transforms into formula (22).

Interchanging in (23) the functions F and G and subtracting the obtained formula from the original one we arrive at the formula

$$(24)\quad \iiint\limits_{D} (F\cdot\Delta G - G\cdot\Delta F)\,dV = \mathop{\iint\!\!\!\!\!\!\bigcirc}\limits_{\partial D}\left(F\frac{\partial G}{\partial n} - G\frac{\partial F}{\partial n}\right)d\sigma$$

known as the **second Green's formula.**

For the velocity field of a fluid the integral on the right of (21) is equal to the quantity of the fluid flowing out of the domain D. Hence, the positiveness of the integral indicates the presence of *sources* in the domain, i.e. points where the fluid appears, and its being negative indicates the presence of *sinks*, i.e. points where the fluid disappears. (It is convenient, however, to call sinks, too, sources but with a negative yield.) The integral itself expresses the full capacity of the sources of the field in the domain D. Dividing the integral by the volume of the domain, therefore, we obtain the average capacity of these sources in D.

Fixing in the domain D some point M_0, consider a sequence $\{D_n\}$ of domains contracting to that point, i.e. such that $M_0 \in D_n$ and $\operatorname{diam} D_n \to 0$ as $n \to \infty$, where $\operatorname{diam} D_n$ is the diameter of the domain D_n. Then the limit

$$\lim_{n\to\infty}\frac{\iint\limits_{\partial D_n} a_n\,d\sigma}{V_n}$$

(when it exists), V_n being the volume of the domain D_n, is nothing but the *density of the sources* of the field at M_0. But according to (21) this limit is equal to

$$\lim_{n\to\infty}\frac{\iiint\limits_{D_n} \operatorname{div}\mathbf{a}\,dV}{V_n}$$

and hence is equal to the value of the function $\operatorname{div}\mathbf{a}$ at M_0. Thus the *divergence of a vector field is nothing but the density of its sources:*

$$\operatorname{div}\mathbf{a} = \lim_{n\to\infty}\frac{\iint\limits_{D_n} a_n\,d\sigma}{V_n}.$$

This yields a new definition of divergence (and in particular, for $\mathbf{a} = \operatorname{grad} F$, a new definition of the Laplace operator ΔF) which has the advantage of being a physical visualization (and explains why fields with $\operatorname{div} \mathbf{a} = 0$ are called source-free fields).

Remark 1. It is helpful to keep in mind that this definition of divergence and of the Laplace operator is also suitable for nonsmooth (even disconnected) fields and functions. It is only necessary that there should exist an appropriate limit.

As we know (see Example 1 above) the field

$$\mathbf{a} = -\frac{\mathbf{r}}{r^3} m_0 \tag{25}$$

of gravitating mass has no sources (in the domain $\mathbb{R}^2 \setminus \{0\}$, where the field is defined). On the other hand, if S_ε is a sphere of radius ε with centre at a point $\mathbf{0}$, then

$$\oint\!\!\!\oint_{S_\varepsilon} a_n \, d\sigma = -m_0 \oint\!\!\!\oint_{S_\varepsilon} \frac{d\sigma}{r^2} = -\frac{m_0}{\varepsilon^2} \oint\!\!\!\oint_{S_\varepsilon} d\sigma = -4\pi m_0 \tag{26}$$

since the integral $\oint\!\!\!\oint_{S_\varepsilon} d\sigma$ is equal to the area $4\pi\varepsilon^2$ of the sphere S_ε. This means that *the source of the gravitational field* (25) *is at the point* $\mathbf{0}$ *and that its capacity is proportional to the mass.* (The minus sign in (26) indicates that in fact it is a sink rather than a source.)

Thus one can say that the sources (sinks) of gravitational fields are gravitating masses.

Remark 2. By introducing Newtonian potentials of massive bodies into consideration it is possible to show that the conclusion remains valid also for gravitational fields of any configuration, but all this goes far beyond the scope of our exposition. [This kind of questions are dealt with by **potential theory** presented in the course in equations of mathematical physics.]

The surface integral appearing in the Stokes formula (see formula (14) of Lecture 27) is by virtue of the general formula (20) nothing else but the flux

$$\iint\limits_{\mathscr{D}} (\operatorname{curl} \mathbf{a})_n \, d\sigma$$

of the curl of a field curl **a** over an oriented surface $\mathscr{D}$. To interpret in a similar manner the integral over the border set of a surface we first notice that a differential form $\omega = P\,dx + Q\,dy + R\,dz$ corresponding to a vector field $\mathbf{a} = P\mathbf{i} + Q\mathbf{j} + R\mathbf{k}$ may be represented as a scalar product $\mathbf{a}\,d\mathbf{r}$, where

$$d\mathbf{r} = dx \cdot \mathbf{i} + dy \cdot \mathbf{j} + dz \cdot \mathbf{k},$$

and hence the integral of ω over an arbitrary curve γ may be written as

$$\int\limits_{\gamma} \mathbf{a}\,d\mathbf{r}. \tag{27}$$

Let γ be a closed curve.

Definition 2. Integral (27) over a closed curve γ is called the *circulation* of a vector field **a** over γ.

To emphasize that γ is a closed curve integral (27) is in this case designated

$$\oint\limits_{\gamma} \mathbf{a}\,d\mathbf{r}. \tag{28}$$

In particular, circulation (28) is defined over the border set $\partial\mathscr{D}$ of an arbitrary orientable surface $\mathscr{D}$ and is nothing else but the line integral appearing in the Stokes formula.

We thus see that the Stokes formula for surface integrals in $\mathbb{R}^3$ may be rewritten as follows:

$$\iint\limits_{\mathscr{D}} (\operatorname{curl} \mathbf{a})_n \, d\sigma = \oint\limits_{\partial\mathscr{D}} \mathbf{a}\,d\mathbf{r}. \tag{29}$$

In words: *the circulation of a vector field over the border set of a surface is equal to the flux of the curl of that field over the surface.*

For curl **a** there is another integral formula which may be regarded as a version of the Gauss-Ostrogradskii formula.

Let as before $\mathscr{D}$ be an oriented surface in $\mathbb{R}^3$ with a field of unit normal vectors $\mathbf{n}$ giving the orientation of that surface. For any vector field $\mathbf{u} = X\mathbf{i} + Y\mathbf{j} + Z\mathbf{k}$ given on $\mathscr{D}$ we define the *integral of* $\mathbf{u}$ *over* $\mathscr{D}$ as follows:

$$\iint_{\mathscr{D}} \mathbf{u}\, d\sigma = \left(\iint_{\mathscr{D}} X\, d\sigma\right)\mathbf{i} + \left(\iint_{\mathscr{D}} Y\, d\sigma\right)\mathbf{j} + \left(\iint_{\mathscr{D}} Z\, d\sigma\right)\mathbf{k}.$$

In particular, this definition applies to a field $\mathbf{u} = \mathbf{n} \times \mathbf{a}$, where $\mathbf{a}$ is an arbitrary vector field defined in the domain containing the surface $\mathscr{D}$.

Definition 3. In the case where a surface $\mathscr{D}$ is closed (compact and has no border set) the integral

$$\iint_{\mathscr{D}} (\mathbf{n} \times \mathbf{a})\, d\sigma \tag{30}$$

is said to be the *circulation* of a vector field $\mathbf{a}$ over the surface $\mathscr{D}$ and designated

$$\oint\!\!\!\oint_{\mathscr{D}} (\mathbf{n} \times \mathbf{a})\, d\sigma.$$

If $\mathscr{D}$ is an elementary surface and $\mathbf{r} = \mathbf{r}(u, v)$ is its parametrization, then on the grounds we already know

$$\iint_{\mathscr{D}} (\mathbf{n} \times \mathbf{a})\, d\sigma = \iint_{U} ((\mathbf{r}_u \times \mathbf{r}_v) \times \mathbf{a})\, du\, dv, \tag{31}$$

where U is an open set of $\mathbb{R}^2$ on which the parametrization $\mathbf{r} = \mathbf{r}(u, v)$ is given.

But

$$\begin{aligned}
(\mathbf{r}_u \times \mathbf{r}_v) \times \mathbf{a} &= (\mathbf{a}\mathbf{r}_u)\,\mathbf{r}_v - (\mathbf{a}\mathbf{r}_v)\,\mathbf{r}_u \\
&= [(Px_u + Qy_u + Rz_u)\, x_v - (Px_v + Qy_v + Rz_v)\, x_u]\,\mathbf{i} \\
&\quad + [(Px_u + Qy_u + Rz_u)\, y_v - (Px_v + Qy_v + Rz_v)\, y_u]\,\mathbf{j} \\
&\quad + [(Px_u + Qy_u + Rz_u)\, z_v - (Px_v + Qy_v + Rz_v)\, z_u]\,\mathbf{k} \\
&= \left(R\begin{vmatrix} z_u & z_v \\ x_u & x_v \end{vmatrix} - Q\begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix}\right)\mathbf{i} \\
&\quad + \left(P\begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} - R\begin{vmatrix} y_u & y_v \\ z_u & z_v \end{vmatrix}\right)\mathbf{j} \\
&\quad + \left(Q\begin{vmatrix} y_u & y_v \\ z_u & z_v \end{vmatrix} - P\begin{vmatrix} z_u & z_v \\ x_u & x_v \end{vmatrix}\right)\mathbf{k}
\end{aligned}$$

and

$$\iint_{\mathscr{D}} R\,dz\,dx - Q\,dx\,dy$$

$$= \iint_U \left(R \begin{vmatrix} z_u & z_v \\ x_u & x_v \end{vmatrix} - Q \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \right) du\,dv,$$

$$\iint_{\mathscr{D}} P\,dx\,dy - R\,dy\,dz$$

$$= \iint_U \left(P \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} - R \begin{vmatrix} y_u & y_v \\ z_u & z_v \end{vmatrix} \right) du\,dv,$$

$$\iint_{\mathscr{D}} Q\,dy\,dz - P\,dz\,dx$$

$$= \iint_U \left(Q \begin{vmatrix} y_u & y_v \\ z_u & z_v \end{vmatrix} - P \begin{vmatrix} x_u & x_v \\ z_u & z_v \end{vmatrix} \right) du\,dv.$$

Hence

$$\iint_{\mathscr{D}} (\mathbf{n} \times \mathbf{a})\,d\sigma = \left(\iint_{\mathscr{D}} R\,dz\,dx - Q\,dx\,dy \right) \mathbf{i}$$

$$+ \left(\iint_{\mathscr{D}} P\,dx\,dy - R\,dy\,dz \right) \mathbf{j}$$

$$+ \left(\iint_{\mathscr{D}} Q\,dy\,dz - P\,dz\,dx \right) \mathbf{k}.$$

Due to additivity this formula is true for an arbitrary (nonelementary) surface $\mathscr{D}$ as well.

But if $\mathscr{D} = \partial D$, where D is a domain with regular boundary, then by the Gauss-Ostrogradskii formula

$$\oint\!\!\!\oint_{\partial D} R\,dz\,dx - Q\,dx\,dy = \iiint_D \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) dV,$$

$$\oint\!\!\!\oint_{\partial D} P\,dx\,dy - R\,dy\,dz = \iiint_D \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x} \right) dV,$$

$$\oint\!\!\!\oint_{\partial D} Q\,dy\,dz - P\,dz\,dx = \iiint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dV.$$

Since

$$\operatorname{curl}\mathbf{a}=\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)\mathbf{i}+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)\mathbf{j}+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\mathbf{k},$$

this proves that the *circulation of a vector field over the boundary ∂D of a domain D is equal to the integral of the curl of that field over D*

$$(32)\qquad \oiint\limits_{\partial D}(\mathbf{n}\times\mathbf{a})\,d\sigma=\iiint\limits_{D}\operatorname{curl}\mathbf{a}\,dV.$$

[It follows, in particular, that the curl of a vector field may be interpreted as the *density of the circulation* of the field (thus giving, in particular, a new definition of a curl, which is suitable for nonsmooth fields as well).]

Formula (32) is a special case of some general formula whose proof, as is often the case, is substantially shorter than the proof of its special case (32).

Let φ be a linear mapping of $\mathbb{R}^3$ into either a space of functions F or a space of vector fields $\mathfrak{a}$. (The linearity implies that for any vector $\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$ of $\mathbb{R}^3$ there is an equation $\varphi(\mathbf{r})=x\varphi(\mathbf{i})+y\varphi(\mathbf{j})+z\varphi(\mathbf{k})$.) We relate to the mapping φ a function (or field) $\varphi(\nabla)$ defined as follows:

$$\varphi(\nabla)=\frac{\partial}{\partial x}\varphi(\mathbf{i})+\frac{\partial}{\partial y}\varphi(\mathbf{j})+\frac{\partial}{\partial z}\varphi(\mathbf{k}).$$

For example, if $\varphi(\mathbf{r})=\mathbf{r}F$ (the product of a vector $\mathbf{r}$ by a function F), then

$$\varphi(\nabla)=\nabla F=\frac{\partial F}{\partial x}\mathbf{i}+\frac{\partial F}{\partial y}\mathbf{j}+\frac{\partial F}{\partial z}\mathbf{k}=\operatorname{grad}F,$$

and if $\varphi(\mathbf{r})=\mathbf{r}\mathbf{a}$ or $\varphi(\mathbf{r})=\mathbf{r}\times\mathbf{a}$, where $\mathbf{a}$ is a field, then, respectively,

$$\varphi(\nabla)=\nabla\mathbf{a}=\operatorname{div}\mathbf{a}\quad\text{or}\quad\varphi(\nabla)=\nabla\times\mathbf{a}=\operatorname{curl}\mathbf{a}.$$

Let for definiteness $\varphi\colon\mathbb{R}^3\to\mathsf{F}$. Then $\varphi(\nabla)=\operatorname{div}\mathbf{a}$, where $\mathbf{a}=\varphi(\mathbf{i})\mathbf{i}+\varphi(\mathbf{j})\mathbf{j}+\varphi(\mathbf{k})\mathbf{k}$, and hence for any domain D by the Gauss-Ostrogradskii formula (20)

$$\iiint\limits_{D}\varphi(\nabla)\,dV=\oiint\limits_{\partial D}a_n\,d\sigma.$$

But if $\mathbf{n} = \cos\alpha\cdot\mathbf{i} + \cos\beta\cdot\mathbf{j} + \cos\gamma\cdot\mathbf{k}$, then

$$\begin{aligned} a_n = \mathbf{an} &= \varphi(\mathbf{i})\cos\alpha + \varphi(\mathbf{j})\cos\beta + \varphi(\mathbf{k})\cos\gamma \\ &= \varphi(\cos\alpha\cdot\mathbf{i} + \cos\beta\cdot\mathbf{j} + \cos\gamma\cdot\mathbf{k}) = \varphi(\mathbf{n}) \end{aligned}$$

and hence

$$\iiint\limits_{D} \varphi(\nabla)\, dV = \oiint\limits_{\partial D} \varphi(\mathbf{n})\, d\sigma. \tag{33}$$

If $\varphi\colon \mathbb{R}^3 \to \alpha$, then applying (33) to each component of the mapping φ we immediately see that the formula is true in this case too.

Formula (33) is called the **generalized Gauss-Ostrogradskii formula.**

When $\varphi(\mathbf{r}) = \mathbf{ra}$ it goes over into formula (20) and when $\varphi(\mathbf{r}) = \mathbf{r} \times \mathbf{a}$, into formula (32).

Lecture 29

Periods of differential forms · Singular simplexes, chains, cycles, and boundaries · Stokes' theorem for chain integrals · Singular homology groups · The de Rham theorem · Cohomology groups of a chain complex · Singular cohomology groups

---

The application of Stokes' theorem to compute cohomology groups $H^m\mathscr{X}$ of a smooth Hausdorff manifold $\mathscr{X}$ (not necessarily without border set) rests on the fact that for any closed differential form $\omega \in \Omega^m\mathscr{X}$ and an arbitrary oriented closed (i.e. compact and without a border set) m-dimensional submanifold $\mathscr{Y} \subset \mathscr{X}$ the integral

$$(1) \qquad I_{\mathscr{Y}}[\omega] = \int_{\mathscr{Y}} \omega$$

depends only on the cohomology class $[\omega] \in H^m\mathscr{X}$ of that form (since if the form ω is exact, then by formula (4) of Lecture 27 this integral is zero). Therefore formula (1) correctly defines some (obviously linear) mapping

$$(2) \qquad I_{\mathscr{Y}}\colon H^m\mathscr{X} \to \mathbb{R}, \quad [\omega] \mapsto I_{\mathscr{Y}}[\omega]$$

of a group $H^m\mathscr{X}$ in $\mathbb{R}$ (i.e. a linear functional on $H^m\mathscr{X}$).

The number $I_{\mathscr{Y}}[\omega]$ is called the *period* of the form ω (or of a cohomology class $[\omega]$) in a submanifold $\mathscr{Y}$.

Of course, if for a cohomology class $[\omega]$ there is a submanifold $\mathscr{Y}$ such that $I_{\mathscr{Y}}[\omega] \neq 0$, then $[\omega] \neq 0$. It turns out—a very difficult theorem!—that conversely *if* $[\omega] \neq 0$, *then there is a submanifold* $\mathscr{Y} \subset \mathscr{X}$ *such that* $I_{\mathscr{Y}}[\omega] \neq 0$.

This theorem is difficult because in constructing a submanifold $\mathcal{Y}$ from a form ω it is a very complicated matter to ensure all the requirements that submanifold should satisfy (for example, the absence of self-intersections). It is a priori clear that the proof will become easier if these requirements are relaxed, i.e. if the class of submanifolds $\mathcal{Y}$ (or, more exactly, functionals (2)) is widened.

The first thing that occurs to one here is to replace the current submanifolds $\mathcal{Y}$ by closed *singular submanifolds*, i.e. by smooth mappings of the form $\gamma\colon \mathcal{D} \to \mathcal{X}$, where $\mathcal{D}$ is an arbitrary m-dimensional oriented compact manifold without border set. Another idea is to consider arbitrary linear combinations of functionals (2). Combining these two considerations results in functionals of the form

$$[\omega] \mapsto \sum_{i=1}^{N} a_i \int_{\gamma_i} \omega, \quad [\omega] \in H^m\mathcal{X},$$

where γ_i are smooth mappings of the form $\mathcal{D}_i \to \mathcal{X}$ and a_i are arbitrary real numbers, each manifold $\mathcal{D}_i$ being closed. For obvious technical reasons, however, it is convenient to introduce into consideration formal linear combinations of the form

$$\gamma = \sum_{i=1}^{N} a_i\gamma_i \tag{3}$$

and assume by definition that

$$\int_{\gamma} \omega = \sum_{i=1}^{N} a_i \int_{\gamma_i} \omega.$$

It may be expected that for a cohomology class $[\omega] \neq 0$ it is easier to construct a linear combination γ for which $\int_{\gamma} \omega \neq 0$ than to find a submanifold $\mathcal{Y}$ with $\int_{\mathcal{Y}} \omega \neq 0$. (Once such a combination γ is found it is possible, if necessary, to raise the question of seeking $\mathcal{Y}$.)

This expectation is in fact justified but, unfortunately, not to the extent we would like, for the problem remains all the same a very difficult one to tackle. Therefore some new idea is necessary here.

Such an idea was suggested nearly eighty years ago by the great French mathematician H. Poincaré, and it is from that time that modern algebraic topology dates.

Poincaré suggested that linear combinations (3) of singular m-dimensional manifolds $\gamma_i\colon \mathscr{D}_i \to \mathscr{X}$ should be considered for which manifolds $\mathscr{D}_i$, while being assumed as before to be compact (so that all integrals should exist), might have a border set. For every such linear combination γ its *boundary* $\partial\gamma$ is defined as follows:

$$\partial\gamma = \sum_{i=1}^{N} a_i\,\partial\gamma_i,$$

where $\partial\gamma_i = \gamma_i\,|_{\partial\mathscr{D}_i}$. (Of course, if all manifolds $\mathscr{D}_i$ are closed, then $\partial\gamma = 0$, but $\partial\gamma = 0$ may hold even when manifolds $\mathscr{D}_i$ have a border set.) It is clear, because of the linearity, that *the Stokes formula remains valid in this case too*, i.e. for any form ω of degree $m - 1$ on a manifold $\mathscr{X}$ we have

$$\int_{\gamma} d\omega = \int_{\partial\gamma} \omega. \tag{4}$$

Therefore if $\partial\gamma = 0$, then the formula

$$I_\gamma[\omega] = \int_{\gamma} \omega, \quad [\omega] \in H^m\mathscr{X},$$

correctly defines some linear functional

$$I_\gamma\colon H^m\mathscr{X} \to \mathbb{R}$$

and the theorem says that for any nonzero cohomology class $[\omega]$ there is a linear combination (3) (with $\partial\gamma = 0$) such that $I_\gamma[\omega] \neq 0$.

It turns out that thus stated, the theorem is comparatively easy to prove. The reason for this is that it is possible to find the corresponding singular manifolds $\gamma_i\colon \mathscr{D}_i \to \mathscr{X}$ in the class of very simple manifolds for which the manifolds $\mathscr{D}_i$ are diffeomorphic to a ball. [Intuitively this is quite clear: every manifold $\mathscr{D}_i$ can be cut into elementary parts diffeomorphic to a ball, and the integral of an arbitrary form over $\mathscr{D}_i$ is equal to the sum of the integrals over those parts.] But if this is the case, then we can take a further step, also proposed by Poincaré, and consider from the very outset only singular manifolds $\mathscr{D} \to \mathscr{X}$ for which the manifold $\mathscr{D}$ is a ball. The corresponding constructions need only to be integrated over the balls, and therefore they can be carried out anew at the elementary level regardless of the general theory of integration over manifolds.

It is more convenient to consider cubes or simplexes rather than balls (neither cubes nor simplexes are manifolds with border set, of course, but points at which there are inflections make up a null set and therefore do not affect integrals). We choose simplexes, mainly by tradition (although reducing multiple integrals to iterated ones is easier for cubes, of course).

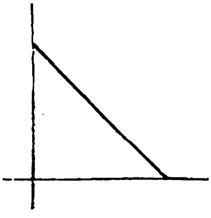
Definition 1. By *a standard m-dimensional simplex* Δ^m is meant a subset of the space $\mathbb{R}^{m+1}$ consisting of points $\mathbf{t} = (t_0, t_1, \ldots, t_m)$ for which

$$0 \leqslant t_0 \leqslant 1, \quad 0 \leqslant t_1 \leqslant 1, \quad \ldots, \quad 0 \leqslant t_m \leqslant 1$$

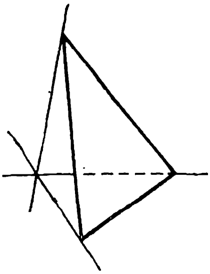
and

$$t_0 + t_1 + \ldots + t_m = 1.$$

[This is a Euclidean simplex in the sense of Lecture 21



Standard simplex Δ^1



Standard simplex Δ^2

with vertices at the end points of the unit vectors $\mathbf{e}_0$, $\mathbf{e}_1, \ldots, \mathbf{e}_m$ of the standard basis for $\mathbb{R}^{m+1}$.]

For $0 \leqslant i \leqslant m$ (and $m > 0$) a mapping

$$\delta_i\colon \Delta^{m-1} \to \Delta^m,$$

$(t_0, t_1, \ldots, t_{m-1}) \mapsto (t_0, t_1, \ldots, t_{i-1}, 0, t_i, \ldots, t_{m-1})$ is defined whose image (generally called the *ith face* of a simplex Δ^m) consists of all points $\mathbf{t} \in \Delta^m$ for which $t_i = 0$.

Each point of Δ^m is uniquely determined by its coordinates $t_1, \ldots, t_m$, which allows this simplex to be identified with a subset of $\mathbb{R}^m$ consisting of points $\mathbf{t} = (t_1, \ldots, t_m)$ such that

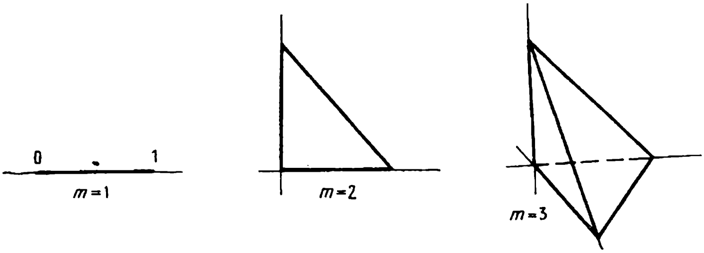
$$0 \leqslant t_1 \leqslant 1, \ldots, 0 \leqslant t_m \leqslant 1$$

and

$$0 \leqslant t_1 + \ldots + t_m \leqslant 1$$

(i.e. with a simplex of $\mathbb{R}^m$ with vertices at $\mathbf{0}$, $\mathbf{e}_1, \ldots, \mathbf{e}_m$).

A mapping $K \to \mathscr{X}$ of an arbitrary set $K \subset \mathbb{R}^m$ into a smooth manifold $\mathscr{X}$ is said to be *smooth* if it is a restriction to K of some smooth mapping $U \to \mathscr{X}$, where U is some subset of $\mathbb{R}^m$ containing the set K.



Standard simplexes projected in $\mathbb{R}^m$

In particular, we may speak of smooth mappings $\Delta^m \to \mathscr{X}$ of a simplex Δ^m into $\mathscr{X}$.

Definition 2. By *a singular m-dimensional simplex* of a smooth manifold $\mathscr{X}$ is meant an arbitrary smooth mapping $\sigma\colon \Delta^m \to \mathscr{X}$. By its *ith face* $\partial_i\sigma$, $0 \leqslant i \leqslant m$, is meant a mapping (also obviously smooth)

$$\sigma \circ \delta_i\colon \Delta^{m-1} \to \mathscr{X}.$$

The set of all singular m-dimensional simplexes of $\mathscr{X}$ is designated $S_m\mathscr{X}$.

Let G be an arbitrary Abelian group (in additive notation).

Definition 3. By a *singular m-dimensional chain of a manifold $\mathscr{X}$ over a group G* (or with *coefficients in G*) is meant an arbitrary finite formal linear combination

$$\gamma = \sum_{\sigma \in S_m \mathscr{X}} a_\sigma \sigma, \quad a_\sigma \in G, \tag{5}$$

of singular simplexes $\sigma \in S_m\mathscr{X}$ with coefficients in G. (The finiteness of sum (5) means that only a finite number of coefficients a_σ are nonzero.) Such chains form an additive group designated $C_m(\mathscr{X}; G)$. For $G = \mathbb{Z}$ this is nothing else but a free Abelian group generated by a set $S_m\mathscr{X}$, and for $G = \mathbb{R}$ it is a vector space with basis $S_m\mathscr{X}$.

We shall also consider *infinite* linear combinations (5) with a property (known as the **local-finiteness property**, cf. Definition 1 of Lecture 22) such that for any point $p \in \mathscr{X}$, and therefore for any compact set $C \subset \mathscr{X}$, there is its neighbourhood U such that the set of singular simplexes $\sigma \in S_m\mathscr{X}$, for which both $a_\sigma \neq 0$ and $\sigma(\Delta^m) \cap U \neq \varnothing$, is finite. Such linear combinations are called *infinite chains*. They form a group $C_m^{\inf}(\mathscr{X}; G)$ containing $C_m(\mathscr{X}; G)$ as a subgroup.

Clearly the *equation $C_m(\mathscr{X}; G) = C_m^{\inf}(\mathscr{X}; G)$ holds if and only if $\mathscr{X}$ is a compact manifold.*

For every $m \geqslant 1$ we define homomorphisms

$$\partial\colon C_m(\mathscr{X}; G) \to C_{m-1}(\mathscr{X}; G)$$

and

$$\partial\colon C_m^{\inf}(\mathscr{X}; G) \to C_{m-1}^{\inf}(\mathscr{X}; G),$$

assuming for any finite (or, respectively, infinite) chain (5)

$$\partial\gamma = \sum_{\sigma \in S_m\mathscr{X}} a_\sigma\, \partial\sigma, \tag{6}$$

where

$$\partial\sigma = \sum_{i=0}^{m} (-1)^i\, \partial_i\sigma,$$

i.e. as follows:

$$\partial\gamma = \sum_{\tau \in S_{m-1}\mathscr{X}} \left(\sum_{i=0}^{m} \sum_{\substack{\sigma \in S_m\mathscr{X}, \\ \partial_i\sigma=\tau}} (-1)^i a_\sigma \right) \tau,$$

where the inner sum is taken over all simplexes $\sigma \in S_m\mathscr{X}$ for which $\partial_i\sigma = \tau$ (the set of such simplexes with $a_\sigma \neq 0$ is finite, even if the chain γ is infinite, and therefore this sum makes sense).

Definition 4. A chain $\partial\gamma$ is called the *boundary* of a chain γ. Chains for which $\partial\gamma = 0$ are called *cycles.*

For any differential form ω of degree m on a manifold $\mathscr{X}$ and any singular simplex $\sigma: \Delta^m \to \mathscr{X}$ a form $\sigma^*\omega$ is defined on Δ^m (or, more exactly, on some open set $U \supset \Delta^m$). In coordinates $t_1, \ldots, t_m$ this form is as follows:

$$\sigma^*\omega = w\, dt_1 \wedge \ldots \wedge dt_m,$$

where $w = w(\mathbf{t})$ is some smooth function. By definition we put

$$(7) \qquad \int_\sigma \omega = \int \ldots \int_{\Delta^m} w(\mathbf{t})\, dt.$$

In other words,

$$(8) \qquad \int_\sigma \omega = \int_0^1 dt_1 \int_0^{s_1} dt_2 \ldots \int_0^{s_m} w(\mathbf{t})\, dt_m,$$

where $s_1 = 1 - t_1, \ldots, s_m = 1 - t_1 - \ldots - t_m$.

The integral of ω over an arbitrary m-dimensional chain γ with coefficients in the field $\mathbb{R}$ is defined by additivity: if $\gamma = \sum a_\sigma\sigma$, then

$$(9) \qquad \int_\gamma \omega = \sum_{\sigma \in S_m\mathscr{X}} a_\sigma \int_\sigma \omega.$$

If the chain γ is infinite, then for sum (9) to make sense (to be finite), we require that the form ω should be finite.

Theorem 1 (Stokes' theorem for chain integrals). *For any m-dimensional singular chain γ of a smooth manifold $\mathscr{X}$ with coefficients in the field $\mathbb{R}$ and any (finite, if the chain γ is infinite) differential form ω of degree $m - 1$ there is an equation*

$$\int_\gamma d\omega = \int_{\partial\gamma} \omega. \tag{10}$$

[This equation is of course a special case of the general equation (4), but we give here its direct proof based only on formulas (6), (7), and (9).]

Proof. It follows directly from formulas (6) and (9) that it suffices to prove equation (10) only for the case where the chain γ is a simplex σ. It may be assumed without loss of generality that $\sigma^*\omega$ is of the form

$$\sigma^*\omega = w\, dt_1 \wedge \dots \wedge \widehat{dt_k} \wedge \dots \wedge dt_m,$$

where $0 \leqslant k \leqslant m$ (and $\hat{}$ as ever indicates that the corresponding factor must be omitted) and that hence $\sigma^*(d\omega) = d(\sigma^*\omega)$ is of the form

$$\sigma^*(d\omega) = (-1)^{k-1} \frac{\partial w}{\partial t_k} dt_1 \wedge \dots \wedge dt_m.$$

Then

$$\int_\sigma d\omega = \int \underset{\Delta^m}{\dots} \int (-1)^{k-1} \frac{\partial w}{\partial t_k}\, dt$$

$$= (-1)^{k-1} \int \underset{\Delta^{m-1}_{(k)}}{\dots} \int \left(\int_0^{1-t} \frac{\partial w}{\partial t_k}\, dt_k \right) dt_{(k)},$$

where $t = t_1 + \dots + \hat{t}_k + \dots + t_m$, $dt_{(k)} = dt_1 \dots \widehat{dt_k} \dots dt_m$ and $\Delta^{m-1}_{(k)}$ is a simplex Δ^{m-1} in $\mathbb{R}^{m-1}$ with coordinates $t_1, \dots, \hat{t}_k, \dots, t_m$. Since

$$\int_0^{1-t} \frac{\partial w}{\partial t_k}\, dt_k = w \Big|_{t_k=0}^{t_k=1-t} = w \Big|_{t_k=1-t} - w \Big|_{t_k=0},$$

where $w\,|_{t_k=r}$ is a restriction of a function w to the hyperplane $t_k = r$ (considered as à function of $t_1, \ldots, \hat{t}_k, \ldots, t_m$) this proves that

$$(11)\quad \int_\sigma d\omega = (-1)^{k-1} \int \cdots \int_{\Delta^{m-1}_{(k)}} (w\,|_{t_k=1-t} - w|_{t_k=0})\, dt_{(k)}.$$

On the other hand, since $t_i \circ \delta_i = 0$ (and hence $\delta_i^* (dt_i) = 0$), we have $(\partial_i \sigma)^*\, \omega = \delta_i^* (\sigma^* \omega) = 0$ for $i \neq 0$, k, and therefore

$$\int_{\partial\sigma} \omega = \int_{\partial_0 \sigma} \omega + (-1)^k \int_{\partial_k \sigma} \omega.$$

At the same time, since

$$t_i \circ \delta_k = \begin{cases} t_i & \text{if } i < k, \\ 0 & \text{if } i = k, \\ t_{i-1} & \text{if } i > k, \end{cases}$$

where $t_1, \ldots, t_m$ on the left are coordinates in $\mathbb{R}^m$, and $t_1, \ldots, t_{m-1}$ on the right are coordinates in $\mathbb{R}^{m-1}$, we have

$$(\partial_k \sigma)^*\, \omega = \delta_k^* (\sigma^* \omega) = (w \circ \delta_k)\, dt_1 \wedge \cdots \wedge dt_{m-1},$$

where

$$(\omega \circ \delta_k)(t_1, \ldots, t_{m-1}) = w(t_1, \ldots, t_{k-1}, 0, t_k, \ldots, t_{m-1})$$

and hence

$$\int_{\partial_k \sigma} \omega = \int \cdots \int_{\Delta^{m-1}} w(t_1, \ldots, t_{k-1}, 0, \ldots, t_k, \ldots, t_{m-1})\, dt_1 \ldots dt_{m-1}.$$

Denoting the variables $t_k, \ldots, t_{m-1}$ by $t_{k+1}, \ldots, t_m$ we may rewrite this integral as

$$\int_{\partial_k \sigma} \omega = \int \cdots \int_{\Delta^m_{(k)}} (w\,|_{t_k=0})\, dt_{(k)}.$$

Similarly, since $t_i \circ \sigma_0 = t_{i-1}$, $i = 1, \ldots, m$, where $t_0 = 1 - t_1 - \ldots - t_{m-1}$, we have

$$\int_{\partial_0\sigma} \omega = \int \ldots \int_{\Delta^{m-1}} w(1 - t_1 - \ldots - t_{m-1}, \\ t_1, \ldots, t_{m-1})\, dt_1 \ldots dt_{m-1}.$$

On making a change of variables

$$\begin{array}{ll}
t'_1 = 1 - t_1 - \ldots - t_{m-1}, & t_1 = t'_2, \\
t'_2 = t_1, & t_2 = t'_3, \\
\cdots\cdots\cdots\cdots\cdots\cdots & \\
t'_{k-1} = t_{k-2}, & t_{k-2} = t'_{k-1}, \\
t'_{k+1} = t_k, & t_{k-1} = 1 - t'_1 \\
 & \qquad - \ldots - \hat{t}'_k - \ldots - t'_m, \\
t'_{k+2} = t_{k+1}, & t_k = t'_{k+1}, \\
\cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots \\
t'_m = t_{m-1}, & t_{m-1} = t'_m
\end{array}$$

(whose Jacobian is equal to $(-1)^{k-1}$) and dropping the primes we get

$$\int_{\partial_0\sigma} \omega = (-1)^{k-1} \int \ldots \int_{\Delta^{m-1}_{(k)}} (w\,|_{t_k = 1 - t})\, dt_{(k)}.$$

Therefore

$$\int_{\partial\sigma} \omega = (-1)^{k-1} \int \ldots \int_{\Delta^{m-1}_{(k)}} (w\,|_{t_k = 1 - t} - w\,|_{t_k = 0})\, dt_{(k)}. \tag{12}$$

A comparison of formulas (11) and (12) proves the theorem. ☐

It follows from Theorem 1 that for any cycle γ and integral

$$I_\gamma \omega = \int_\gamma \omega$$

of a closed form ω over γ depends only on the cohomology class $[\omega] \in H^m\mathscr{X}$ of that form, i.e. the formula

$$(13) \qquad I_\gamma[\omega] = \int_\gamma \omega$$

correctly defines some homomorphism (linear functional)

$$(14) \qquad I_\gamma\colon H^m\mathscr{X} \to \mathbb{R}.$$

In a similar manner, for any infinite cycle γ the same formula (14), but with a finite form ω, correctly defines a linear functional

$$(15) \qquad I_\gamma\colon H^m_{\text{fin}}\mathscr{X} \to \mathbb{R}.$$

In both cases the number $I_\gamma[\omega]$ is called the *period* of the form ω (or of the cohomology class $[\omega]$) in the cycle γ.

We now study the dependence of the function I_γ on the cycle γ. To do this we must first develop the appropriate algebraic tools.

Lemma 1. *If* $0 \leqslant i < j \leqslant m$, *then*

$$(16) \qquad \partial_i\partial_j\sigma = \partial_{j-1}\partial_i\sigma$$

for any m-dimensional singular simplex σ *of a manifold* $\mathscr{X}$.

Proof. Since $\partial_i\partial_j\sigma = \sigma \circ \delta_j \circ \delta_i$ and $\partial_{j-1}\partial_i\sigma = \sigma \circ \delta_i \circ \delta_{j-1}$, it suffices to prove that

$$\delta_j \circ \delta_i = \delta_i \circ \delta_{j-1}, \quad 0 \leqslant i < j \leqslant m.$$

But by definition the action of the mapping δ_i is to insert a zero into the ith place in the vector $\mathbf{t} = (t_0, \ldots, t_m)$. Since after the insertion the number $j > i$ is acquired by the place which had the number $j - 1$, the mapping $\delta_i \circ \delta_{j-1}$ inserts zeroes into the ith and the jth places. On the other hand, after applying the mapping δ_j all places with numbers less than j retain their numbers, and therefore the mapping $\delta_j \circ \delta_i$ also inserts zeroes into the ith and jth places. Hence $\delta_j \circ \delta_i = \delta_i \circ \delta_{j-1}$. □

Proposition 1. *For any m-dimensional* ($m \geqslant 2$) *chain* γ *of a manifold* $\mathscr{X}$ *over an arbitrary group G there is an equation*

$$(17) \qquad \partial\partial\gamma = 0.$$

Proof. (Cf. the proof of Proposition 1 of Lecture 21.) It is obviously sufficient to prove (17) for the case where the chain γ is a singular simplex $\sigma \in S_m\mathscr{X}$. But for any simplex σ the chain $\partial\partial\sigma$ is obviously the sum of chains of the form $(-1)^{i+j}\,\partial_i\partial_j\sigma$, where $0 \leqslant i \leqslant m-1$ and $0 \leqslant j \leqslant m$. On the other hand, by Lemma 1 if $i < j$, then $\partial_i\partial_j\sigma = \partial_{j-1}\partial_i\sigma$. For any pair (i, j), with $i < j$, there are therefore two equal terms in the chain $\partial\partial\sigma$, one with the sign $(-1)^{i+j}$ and the other with the opposite sign $(-1)^{(j-1)+i}$. All the terms of the chain $\partial\partial\sigma$ therefore cancel out. □

Proposition 1 implies that for any $m \geqslant 1$ the *boundary group* $B_m(\mathscr{X}; G)$ (the image of a homomorphism $\partial\colon C_{m+1}(\mathscr{X}; G) \to C_m(\mathscr{X}; G)$) is contained in the *cycle group* $Z_m(\mathscr{X}; G)$ (in the kernel of a homomorphism $\partial\colon C_m(\mathscr{X}; G) \to C_{m-1}(\mathscr{X}; G)$), and similarly the *infinite boundary group* $B_m^{\inf}(\mathscr{X}; G)$ (the image of a homomorphism $\partial\colon C_{m+1}^{\inf}(\mathscr{X}; G) \to C_m^{\inf}(\mathscr{X}, G)$) is contained in the *infinite cycle group* $Z_m^{\inf}(\mathscr{X}; G)$ (the kernel of a homomorphism $\partial\colon C_m^{\inf}(\mathscr{X}; G) \to C_{m-1}^{\inf}(\mathscr{X}; G)$). The quotient groups

$$H_m(\mathscr{X}; G) = Z_m(\mathscr{X}; G)/B_m(\mathscr{X}; G)$$

and

$$H_m^{\inf}(\mathscr{X}; G) = Z_m^{\inf}(\mathscr{X}; G)/B_m^{\inf}(\mathscr{X}; G)$$

are therefore defined, which are called an m-dimensional (or the mth) *group of singular homologies* (respectively, *singular infinite homologies*) *of a manifold* $\mathscr{X}$ *with coefficients in a group* G (*or over a group* G).

Remark 1. Homology groups are also called *homology groups with compact supports* and infinite-homology groups are also called *homology groups with arbitrary supports.*

Remark 2. The construction can be carried over word for word to any topological spaces $\mathscr{X}$ (it is only necessary to require continuity instead of smoothness in the definition of a singular simplex). Thus for a smooth manifold, along with the above homology groups (said to be based on smooth singular simplexes) there also arise *topological homology groups* (based on any continuous singular simplexes). However, it is possible to show, by a conceptually simple but somewhat cumbersome construction, that a *natural mapping of the latter groups into the former is an isomorphism.*

Elements of homology groups (classes of cycles modulo boundaries) are called *homology classes*, and two cycles belonging to the same homology class (i.e. differing by a boundary) are called *homologous.* Accordingly boundaries are also called *cycles homologous to zero.*

Now let the group G be again the field $\mathbb{R}$.

As we know, for a cycle γ and a closed form ω the integral

$$\int_{\gamma} \omega$$

depends only on the cohomology class $\mathbf{x} = [\omega]$ of ω. From Theorem 1 it now follows that this integral only depends on the homology class ξ of γ (since

$$\int_{\partial\beta} \omega = \int_{\beta} d\omega = 0$$

for any chain β). Therefore the formula

$$\langle \xi, \mathbf{x} \rangle = \int_{\gamma} \omega \tag{18}$$

correctly defines some pairing (see Lecture II.4) between vector spaces $H_m(\mathscr{X}; \mathbb{R})$ and $H^m\mathscr{X}$ (and between vector spaces $H_m^{\mathrm{inf}}(\mathscr{X}; \mathbb{R})$ and $H_{\mathrm{fin}}^m \mathscr{X}$).

Recall (ibid.) that a pairing between vector spaces $\mathscr{V}$ and $\mathscr{W}$ is said to be *nondegenerate* if for any nonzero vector $\xi \in \mathscr{V}$ there is a (trivially nonzero) vector $\mathbf{x} \in \mathscr{W}$ such that $\langle \xi, \mathbf{x} \rangle \neq 0$, and vice versa.

Theorem 2. *For any oriented paracompact and Hausdorff manifold $\mathscr{X}$ both pairings* (18) *(between $H_m(\mathscr{X}; \mathbb{R})$ and $H^m\mathscr{X}$ and between $H_m^{\mathrm{inf}}(\mathscr{X}; \mathbb{R})$ and $H_{\mathrm{fin}}^m \mathscr{X}$) are nondegenerate.*

Let us discuss this theorem at greater length. (We shall not prove it.)

For every pairing between vector spaces $\mathscr{V}$ and $\mathscr{W}$ (over some field $\mathbb{K}$) and any vector $\xi \in \mathscr{V}$ the formula

$$I_\xi \mathbf{x} = \langle \xi, \mathbf{x} \rangle, \quad \mathbf{x} \in \mathscr{W},$$

defines on $\mathscr{W}$ a linear functional $I_\xi \colon \mathscr{W} \to \mathbb{K}$, i.e. a vector of the conjugate space $\mathscr{W}' = \operatorname{Hom}(\mathscr{W}, \mathbb{K})$. The nondegeneracy of a pairing in $\mathbf{x}$ implies that for any nonzero element $\mathbf{x} \in \mathscr{W}$ there is an element $\xi \in \mathscr{V}$ such that $I_\xi \mathbf{x} \neq 0$.

Similarly, for any element $\mathbf{x} \in \mathscr{W}$ the formula $I_x \xi = \langle \xi, \mathbf{x} \rangle$ defines a functional $I_x \colon \mathscr{V} \to \mathbb{K}$, and the nondegeneracy of a pairing in ξ implies that for any nonzero element $\xi \in \mathscr{W}$ there is an element $\mathbf{x} \in \mathscr{V}$ such that $I_x \xi \neq 0$.

For pairing (18) the functional I_ξ is precisely the functional I_γ (where γ is a cycle such that $\xi = [\gamma]$). Thus Theorem 2 states that for any closed form ω not cohomologous to zero there is a cycle γ (trivially not homologous to zero) such that $I_\gamma \omega \neq 0$. This is precisely the statement we started this lecture with.

But Theorem 2 also states that, conversely, for any cycle γ not homologous to zero there is a closed form ω (trivially not cohomologous to zero) such that $I_\omega \gamma \neq 0$ (and hence $I_\gamma \omega \neq 0$).

Moreover, Theorem 2 says that similar statements are true with respect to finite forms and infinite chains. [**Warning.** The two parts of the theorem do not, contrary to what might be thought, reduce directly to each other, because a finite form finitely noncohomologous to zero may nevertheless be cohomologous to zero, and similarly a finite chain nonhomologous to zero may be the boundary of an infinite chain.]

Correspondences $\xi \mapsto I_\xi$ and $\mathbf{x} \mapsto I_x$ obviously define some homomorphisms

$$\mathscr{V} \to \mathscr{W}' \quad \text{and} \quad \mathscr{W} \to \mathscr{V}' \tag{19}$$

of vector spaces $\mathscr{V}$ and $\mathscr{W}$ into the conjugate spaces $\mathscr{W}'$ and $\mathscr{V}'$. The nonsingularity of a pairing is equivalent to the fact that *both homomorphisms* (19) *are monomorphisms.*

For finite-dimensional spaces $\mathscr{V}$ and $\mathscr{W}$, we know (Proposition 4 of Lecture II.4) that for a nonsingular pairing mappings (19) are even isomorphisms. For infinite-dimensional spaces this in general is not so, and, moreover, it may happen that one of the mappings is isomorphic and the other is not.

Example 1. Let S be an infinite set, $\mathscr{V}$ be the vector space of all formal finite sums $\xi = \sum a_\sigma \sigma$, where $\sigma \in S$, $a_\sigma \in \mathbb{K}$, and $\mathscr{W}$ be the vector space of all functions $\mathbf{x}\colon S \to \mathbb{K}$. Then the formula

$$\langle \xi, \mathbf{x} \rangle = \sum_{\sigma \in S} a_\sigma \mathbf{x}(\sigma)$$

correctly defines an (obviously nonsingular) pairing between $\mathscr{V}$ and $\mathscr{W}$. For this pairing, every functional

$$I_\xi(\mathbf{x}) = \sum_{\sigma \in S} a_\sigma \mathbf{x}(\sigma)$$

has the property that $I_\xi \mathbf{x} \neq 0$ only for elements $\mathbf{x} \in \mathscr{W}$ in some finite-dimensional subspace of $\mathscr{W}$ (consisting of functions $\mathbf{x}$ which are nonzero only on those elements $\sigma \in S$ for which $a_\sigma \neq 0$). Hence the mapping $\xi \mapsto I_\xi$ of $\mathscr{V}$ into $\mathscr{W}'$ is trivially not epimorphic. On the contrary, since any functional $\varphi\colon \mathscr{V} \to \mathbb{K}$ can be represented as I_x, $\mathbf{x} \in \mathscr{W}$ (it suffices to put $\mathbf{x}(\sigma) = \varphi(\sigma)$ for any $\sigma \in S$), the mapping $\mathbf{x} \mapsto I_x$ of $\mathscr{W}$ into $\mathscr{V}'$ is an isomorphism.

The following theorem is therefore a refinement of Theorem 2.

Theorem 3. *For any paracompact Hausdorff manifold $\mathscr{X}$ homomorphisms*

$$H^m \mathscr{X} \to H_m(\mathscr{X}; \mathbb{R})', \quad \mathbf{x} \mapsto I_x, \tag{20}$$

and

$$H_m^{\mathrm{inf}}(\mathscr{X}; \mathbb{R}) \to (H^m_{\mathrm{fin}} \mathscr{X})', \quad \xi \mapsto I_\xi, \tag{21}$$

where I_x and I_ξ are functionals $H_m(\mathscr{X}; \mathbb{R}) \to \mathbb{R}$ and $H^m_{\mathrm{fin}} \mathscr{X} \to \mathbb{R}$ defined, respectively, as follows:

$$I_x \xi = \int_\gamma \omega, \quad \mathbf{x} = [\omega], \ \xi = [\gamma],$$

and

$$I_\xi \mathbf{x} = \int_\gamma \omega, \quad \mathbf{x} = [\omega]_{\mathrm{fin}}, \quad \xi = [\gamma]^{\mathrm{inf}},$$

are isomorphisms.

[On the contrary, the "dual" mappings

$$H_m(\mathscr{X};\ \mathbb{R}) \to (H^m\mathscr{X})' \text{ and } H^m_{\mathrm{fin}}\,\mathscr{X} \to H^{\mathrm{inf}}_m(\mathscr{X};\ \mathbb{R})'$$

are not in general isomorphisms.]

Theorem 3 (together with Theorem 2) is known as the **de Rham theorem** (de Rham being the first to prove it for compact manifolds).

Isomorphism (20) reduces computing the group $H^m\mathscr{X}$ to computing the group $H_m(\mathscr{X};\ \mathbb{R})'$, for which special methods, sufficiently effective in principle, have been developed in algebraic topology.

The construction of homology groups $H_m(\mathscr{X};\ G)$ and $H^{\mathrm{inf}}_m(\mathscr{X};\ G)$ splits into two stages, i.e. construction of chain groups $C_m(\mathscr{X};\ G)$ and $C^{\mathrm{inf}}_m(\mathscr{X};\ G)$ together with homomorphisms ∂ (the first stage) and construction of the homology groups themselves from the chain groups (the second stage). It is appropriate, as in a similar situation with cohomology groups, to have a separate second stage.

Definition 5. A family

$$C_\bullet\colon \qquad C_0 \xleftarrow{\partial} C_1 \leftarrow \ldots \leftarrow C_{m-1} \xleftarrow{\partial} C_m \leftarrow \ldots$$

of groups and homomorphisms is said to be a *chain complex* if

$$\partial\partial\gamma = 0$$

for any element $\gamma \in C_m$, $m \geqslant 1$. Elements of C_m are called *m-dimensional chains*; a homomorphism ∂ is called a *boundary operator*; chains γ for which $\partial\gamma = 0$, i.e. which are in the kernel

$$Z_m C_\bullet = \operatorname{Ker}(\partial\colon C_m \to C_{m-1})$$

of a homomorphism $\partial\colon C_m \to C_{m-1}$ are *cycles* (for $m = 0$ it is assumed by convention that $Z_m C_\bullet = C_0$); chains of the form $\partial\gamma$, i.e. chains in the image

$$B_m C_\bullet = \operatorname{Im}(\partial\colon C_{m+1} \to C_m)$$

of a homomorphism $\partial\colon C_{m+1} \to C_m$ are called *boundaries* (or cycles *homologous to zero*); a quotient group

$$H_m C_\bullet = Z_m(C_\bullet)/B_m(C_\bullet)$$

is an *m-dimensional* (or the mth) *homology group* of a complex $C_\bullet$; the elements of the group are called *m-dimensional homology classes*, and cycles which are in the same homology class, i.e. differ by a boundary, are *homologous*.

Groups $H_m(\mathscr{X}; G)$ are thus homology groups $H_m C_\bullet(\mathscr{X}; G)$ of a complex

$$C_\bullet(\mathscr{X}; G) = \{C_m(\mathscr{X}; G); \partial\},$$

and groups $H_m^{\inf}(\mathscr{X}; G)$ are homology groups of a complex

$$C_\bullet^{\inf}(\mathscr{X}; G) = \{C_m^{\inf}(\mathscr{X}; G); \partial\}.$$

In the case where a chain complex $C_\bullet = \{C_m; \partial\}$ consists of vector spaces over a field $\mathbb{K}$, for any $m \geqslant 0$ we can construct a conjugate vector space

$$C^m = (C_m)' = \operatorname{Hom}(C_m; \mathbb{K}).$$

In modern mathematics it is customary to distinguish a passage to a dual (conjugate) situation with a prefix "co" (for example: vectors and covectors). Accordingly elements of a space C^m are called *m-dimensional cochains* of a chain complex $C_\bullet$.

For any cochain $c \in C^m$ the formula

$$(\delta c)(\gamma) = c(\partial\gamma), \quad \gamma \in C_{m+1},$$

defines some cochain δc, and since

$$(\delta\delta c)(\gamma) = c(\partial\partial\gamma) = 0$$

for any chain $\gamma \in C_{m+2}$, we have

$$\delta\delta c = 0$$

for any cochain $c \in C^m$. By definition (see Definition 1 of Lecture 20) this means that a *family* $C^\bullet = \{C^m; \partial\}$ *of groups and homomorphisms is a cochain complex.*

Cocycles $c \in C^m(C^\cdot)$ of the complex are characterized by the relation

$$(\delta c)(\gamma) = c(\partial\gamma) = 0, \quad \gamma \in C_m,$$

i.e. by being zero on a subgroup of boundaries $B_m C_\cdot \subset C_m$. By definition (see Lecture II.4) this means that

$$Z^m C^\cdot = \operatorname{Ann} B_m C_\cdot$$

(*cocycles make up the annihilator of a boundary group*). Similarly, if $\partial\gamma = 0$, then $(\delta e)(\gamma) = e(\partial\gamma) = 0$ for any cochain e and, conversely, if $c(\gamma) = 0$ when $\partial\gamma = 0$, then the formula $e(\partial\gamma) = c(\gamma)$ correctly defines a linear functional $e: B_{m-1}C_\cdot \to \mathbb{K}$ with the property that arbitrarily extended, with linearity preserved, to the entire space C_{m-1}, it gives a cochain $e: C_{m-1} \to \mathbb{K}$ such that $(\delta e)(\gamma) = e(\partial\gamma) = c(\gamma)$, i.e. such that $\delta e = c$. This proves that

$$B^m C^\cdot = \operatorname{Ann} Z_m C_\cdot$$

(*coboundaries make up the annihilator of a cycle group*).

The extendability of the functional e from the subspace $B_{m-1}C^\cdot$ to the entire space C_{m-1} is a general fact: for any subspace $\mathcal{P}$ of an arbitrary—even infinite-dimensional!—vector space $\mathcal{V}$ every linear functional $e: \mathcal{P} \to \mathbb{K}$ allows an extension to $\mathcal{V}$. To prove this it suffices to choose for $\mathcal{P}$ an additional subspace $\mathfrak{Q}$ whose existence is proved in Lecture 22 in connection with Remark 3, and to give e arbitrarily on $\mathfrak{Q}$.

For cohomology groups $H^m(C^\cdot)$ of a complex $C^\cdot$ this implies that for any cohomology class $x \in H^m C^\cdot$ the formula

$$(\#x)(\xi) = c(\gamma), \quad \xi \in H_m C_\cdot,$$

where c is an arbitrary cocycle of a cohomology class x and γ is an arbitrary cycle of a homology class ξ correctly gives a functional

$$\#x: H_m C_\cdot \to \mathbb{K}$$

and that the resulting mapping

$$\#: H^m C^\cdot \to (H_m C_\cdot)', \quad x \mapsto \#x,$$

is an isomorphism. [That this mapping is monomorphic follows from the equation $B^m C^\cdot = \operatorname{Ann} Z_m C_\cdot$ and that it is epimorphic is ensured by the possibility of any linear functional $Z_m C_\bullet \to \mathbb{K}$ being extended to the entire space C_m.] Thus, *for any chain complex* $C_\bullet$ *consisting of vector spaces and any* $m \geqslant 0$ *there is an isomorphism*

$$\#: H^m C^\bullet \to (H_m C_\bullet)'.$$

This reduces computing cohomology groups to computing homology groups.

For a complex $C_\bullet = C_\bullet(\mathscr{X}; \mathbb{K})$ of singular chains of a smooth manifold $\mathscr{X}$ over a field $\mathbb{K}$ a cochain complex $C^\bullet$ is designated $C^\bullet(\mathscr{X}; \mathbb{K})$. Since for any $m \geqslant 0$ the set $S_m\mathscr{X}$ is a basis for the space $C_m(\mathscr{X}; \mathbb{K})$ and since linear functionals on an arbitrary vector space can naturally be identified with $\mathbb{K}$-valued functions given on its basis, for any $m \geqslant 0$ the component $C^m(\mathscr{X}; \mathbb{K}) = C_m(\mathscr{X}; \mathbb{K})'$ of a complex $C^\bullet(\mathscr{X}; \mathbb{K})$ can be identified with the vector space of all possible functions $S_m\mathscr{X} \to \mathbb{K}$. This in particular allows a group $C^m(\mathscr{X}; G)$ to be defined for an **arbitrary** group G. By definition elements of that group called *m-dimensional singular cochains of a manifold* $\mathscr{X}$ *over the group* G are G-valued functions $c: S_m\mathscr{X} \to G$ defined on the set $S_m\mathscr{X}$ of all m-dimensional singular simplexes of $\mathscr{X}$. The coboundary operator δ is defined by the formula

$$(\delta c)(\sigma) = \sum_{i=0}^{m} (-1)^i c(\partial_i \sigma)$$

which deciphers the formula $(\delta c)(\sigma) = c(\partial c)$.

Cocycles, coboundaries, cohomology classes of the complex $C^\bullet(\mathscr{X}; G)$ are called *singular cocycles*, *coboundaries*, and *cohomology classes* of a manifold $\mathscr{X}$ and the corresponding groups are denoted by $Z^m(\mathscr{X}; G)$, $B^m(\mathscr{X}; G)$, and $H^m(\mathscr{X}; G)$. According to the general result proved above, *for any field* $\mathbb{K}$ *there is an isomorphism*

$$H^m(\mathscr{X}; \mathbb{K}) \to H_m(\mathscr{X}; \mathbb{K})'. \tag{22}$$

In the case where $\mathbb{K} = \mathbb{R}$ every form $\omega \in \Omega^m \mathscr{X}$ defines by the formula

$$\bar{\omega}(\gamma) = \int_\gamma \omega, \quad \gamma \in C_m(\mathscr{X}; \mathbb{R}),$$

some cochain $\overline{\omega} \in C^m(\mathscr{X}; \mathbb{R})$, with

$$\overline{d\omega} = \delta\overline{\omega}. \tag{23}$$

[Indeed,

$$\overline{d\omega}\,(\gamma) = \int_{\gamma} d\omega = \int_{\partial\gamma} \omega = \overline{\omega}\,(\partial\gamma) = (\delta\overline{\omega})\,(\gamma)$$

for any chain $\gamma \in C_m(\mathscr{X}; \mathbb{R})$.] It follows from (23) that the mapping $\omega \mapsto \overline{\omega}$ changes cocycles into cocycles and coboundaries into coboundaries. Therefore this mapping induces some homomorphism

$$H^m\mathscr{X} \to H^m(\mathscr{X}; \mathbb{R}). \tag{24}$$

A comparison of the definitions shows at once that the composition of homomorphism (24) and isomorphism (22) is nothing but homomorphism (20) of the de Rham theorem. To prove the theorem therefore (at least with respect to groups $H^m\mathscr{X}$) it suffices to establish that *homomorphism* (24) *is an isomorphism.* Unfortunately, we cannot afford to carry out the corresponding proof here.

Remark 3. The statement that homomorphism (24) is an isomorphism implies that both approaches to the definition of cohomology groups of a manifold (in terms of forms and singular chains) lead in fact to the same result. There are other approaches to these groups as well, such as the Čech-Leray approach of Lecture 22 which we know to lead to groups isomorphic to groups $H^m\mathscr{X}$. To prove isomorphism (24) it suffices therefore to establish that *singular cohomology groups are isomorphic to Čech-Leray groups* (which were designated by the same symbol $H^m(\mathscr{X}; \mathbb{R})$ in Lecture 22). Forms are not involved in this formulation, and the formulation makes sense for any topological spaces (for which there are appropriately defined Leray coverings). The question therefore may be raised as to the validity of this statement in the general form (for any topological spaces). The answer turns out to be affirmative for polyhedra (see Remark 2 of Lecture 21) but negative in the general case. A general study of the entire range of these questions is the subject of a broad theory which is a constituent part of general topology. Unfortunately, we were able to touch on this theory only by hints. [For example, a comparison of the definitions of cohomology groups of simplicial schemes and of singular cohomology groups of manifolds, and, more generally, of topological spaces, shows their close resemblance. The elucidation of this resemblance in general terms leads to **simplicial sets** whose theory is both beautiful and profound. On the other hand, construction of a cochain complex from a chain one is a poor example and a shadow of algebraic manipulations on chain complexes which are dealt with in **homological algebra**, and so on and so forth. The interested reader may find further information in the book by Bott and Tu referred to in the Preface, and in many other, rather numerous, textbooks of algebraic topology.]

SUBJECT INDEX

Atlas, 115
    maximal, 117
    of ∂-charts, 457
    oriented, 421
    orienting, 420
Augmented double complex, 369

Base, 431
    for the neighbourhood, 130
    of open sets, 131
Binormal vector, 40

Cardinality, 171
Cauchy convergence, 191
Cauchy-Riemann relations, 215
Chart, 112
    compatible, 113
    interior, 458
    overlapping, 113
    positively oriented, 420
    smooth, 458
Chain rule for holomorphic mappings, 218
Christoffel symbol, 100
Circular helix, 42
Coboundary, 345, 505
Coboundary operator, 362
Cochain complex, 331
Cocycles, 504
    singular, 505
Cohomology classes, 331
    two-dimensional, 351
Cohomology group, 331
    singular, 506
Cohomology groups of a double complex, 367
Cohomology groups of a covering, 367
Component, 204
    connected, 204
Connection coefficient, 100
Coordinate vector field, 278
Covering,
    contractible, 168
    finite, 168
    Leray, 377
    nerve of, 358
Curvature, 36
    Gaussian, 187
    mean, 87
    of the curve, 36
    principal, 87
    total, 87
Curve, 65
    curvature of, 36
    equivalent, 30
    generic, 28
    length of, 33
    non-parametrized, 30
    parametrized, 30
    regular, 30
    simple, 27
    smooth, 29
    support of, 27
Cycle,
    homologous to zero, 499
    of the differential d_r, 400

Derivative,
    directional, 479
    normal, 479
Direct product, 376

Form,
    differential, of degree n, 422
    essential, 429
    exact, 340
    finite, of degree n, 428
    independent of dt, 337

Form
  inessential, 429
  of maximal degree, 423
  normed, 429
  period of, 497
  smoothly dependent on t, 337
Formula(s),
  Frenet's, 39
  Gauss-Ostrogradskii, 455
    generalized, 486
  Green's, 456
    first formula, 479
    second formula, 480
  Newton-Leibniz, 456
  Peterson-Codazzi, 103
  Stokes for surface integrals, 463
  transition, 114
  Weingarten, 100
Frenet's formulas for a plane curve, 39
Frenet's moving basis, 38

Group(s),
  arbitrary, 497
  boundary, 498
  Čech cohomology, 378
  Čech-Leray, 506
  cohomology, 367
  of a covering, 367
  of a double complex, 367
  de Rham cohomology, 331
  family of, 503
  graded Abelian, 386
  homology, 499
    with arbitrary supports, 498
  of singular homologies, 498
  symplectic, 198
Gauss theorem, 103
Gaussian coordinates, 107
Gram determinant, 68

Homological algebra, 507

Immersion, 233

Lemma,
  Lebesgue, 146
  Zorn, 182

Lebesgue number, 147
Lie algebra, 195
Line,
  closed, 18
  half-open, 18
  open, 18
  simple, 17

Manifold(s), 117
  C^r-manifold, 118
  Grassmann, 190
  non-smoothable, 136
  orientable, 420
  parallelizable, 279
  smooth, 117
  Stieffel, 190
  topological, 118
Mapping,
  cochain, 332
  complex analytic, 202
  compact, 492
  continuous bijective, 248
  continuous, 111
  continuous at a point, 137
  Grassmann, 190
  holomorphic, 217
  of maximal rank, 233
  smooth, 111
  smooth at a point, 139
  transition, 114
Matrices, 194
  Jacobian, 217
  J-skew-symmetric, 195
  J-unitary, 199
  pseudo-orthogonal of type (p, q), 199
  real, 194
  symplectic, 198
Matrix Lie group, 195
Monomorphism, 351

Number, 171
  Betti, 333
  countable, 171
  Lebesgue, 147
  ordinal finite, 171
    of the first number class, 171
    of the second number class, 171
    of the third number class, 173

Open ball, 23
Open-closed subset, 204
Open submanifold, 129
Ordinal number, 171
    countable, 171
    finite, 171
    limiting, 172
    ordered, 172
    of the first number class, 171
    of the second number class, 171
    of the third number class, 173

Parameter,
    Beltrami mixed differential parameter of functions, 476
    Beltrami's first differential parameter of a function, 476
    Darboux differential, 476
Path, 203
    constant, 203
    reverse, 203
Parametrization, 52
Parametric equations, 53
Pathwise connectedness, 203
Peterson-Codazzi formulas, 103
Piecewise linear topology, 300
Plane,
    normal, 44
    osculating, 44
    simple, 17
    rectifying, 44
Point,
    end, 18,
    interior, 19, 27
    hyperbolic, 86
    nonsingular, 20
    of a curve, 27
    parabolic, 86
    of rectification, 37
    rational, 24
    simple, 20
    singular, 20
Property of monotonicity of dimension, 149

Separation extension, 164
Set, 18
    centred, 180
    closed, 18
    open, 113
    ordered, 170
    partially ordered, 170
    simplicial, 507
    well ordered, 171
Spectral sequence, 402
    degenerate, 403
Singular m-dimensional chain, 492
Singular m-dimensional complex, 491
Singular submanifold, 488
Submersion, 233
Space,
    affine, 19
    Baire, 254
    bicompact, 146
    compact, 146
    completely normal, 161
    connected, 205
    cotangent, 225
    Euclidean, 65
    Hausdorff, 132
    hereditary normal, 162
    locally compact, 254
    locally Euclidean, 131
    normal, 161
    paracompact, 407
    pathwise connected, 203
    quasi-compact, 146
    ruled, 57
    separable, 132
    topological, 127
    totally disconnected, 206
Spectral sequence, 397
Surface,
    abstract, Riemann, 209
    elementary, 52
    of binomials, 59
    of principal normal, 59
    of revolution, 55
    of tangents, 59
    ruled, 57

Tensor, 274
    components of, 276
    of type (a, b)

- Tensor field, 276
- Theorem,
  - Brouwer on the domain invariance, 459
  - cube separation, 158
  - Dieudonne, 410
  - de Rham for Čech cohomology groups, 379
  - de Rham in the Leray form, 377
  - drum, 156
  - étale mapping, 227
  - Fubini, 256
  - Gauss' 103
  - Heine-Borel, 146
  - inverse function, 31
  - inverse mapping, 81
  - Kuratowski-Zorn, 182
  - Lebesgue, on coverings of a cube, 166
  - on the change of variables, 412
  - on change of local coordinates, 228
- Theorem
  - on the homotopic invariance, of degree, 442
  - on the inverse image of a regular value, 239
  - Stokes for chain integrals, 494
  - Stokes for domains with regular boundaries, 450
  - Stokes for manifolds with border set, 462
  - Stokes for submanifolds, 463
  - Unique existence theorem (UES), 47
  - Zermelo, 173
  - Whitney's 20, 246
- Topology,
  - algebraic, 363
  - combinatorial, 363

- Uryson's function, 248

- Vector of principal normal, 40
- Vector space, 278

Учебное Издание
Михаил Михайлович Постников
ЛЕКЦИИ ПО ГЕОМЕТРИИ
Семестр III
ГЛАДКИЕ МНОГООБРАЗИЯ

ТОМ 2

Лицензия ЛР №063377от 23.05.94г.
Подписано к печати 06.06.94г. Формат 60X88/16
Бумага офсетная. Гарнитура обыкновенная.
Печ. л. 15,5. Тираж 1500 экз. Зак № 529

Издательство "УРСС"
111672, г. Москва, ул. Новокасинская 27/174

Московская тип. 4 Комитета РФ по печати
129041 Москва, Б. Переяславская, 46